INTERNATIONAL
ECONOMIC
REVIEW

June, 1968
Vol. 9, No. 2

ECONOMIC COMPARABILITY OF INFORMATION SYSTEMS*

BY JACOB MARSCHAK AND KOICHI MIYASAWA[1]

1. INTRODUCTION

AN INFORMATION SYSTEM is a set of potential messages to be received by the decision maker. It is characterized by the statistical relation of the messages to the payoff-relevant events, and also by the message cost.[2] Neglecting this cost, the (gross) value of an information system for a given user is the (gross) payoff that he would obtain, on the average, if he would respond to each message by the most appropriate decision. Thus (gross) information value depends not only on the statistical relation between messages and events but also on the payoff function. The latter expresses the user's "tastes" and "technology." The ordering of statistically defined information systems by their values is therefore at most a partial one. This contrasts with the complete ordering of information systems (channels) by their equivocation (a statistical parameter used in the classical information theory that disregards variation of payoff functions from user to user).

Indeed, if "noise" is defined to increase with equivocation[3] a "noisy" information system may be more valuable to a given user than a noiseless one: the betting sports fan may have reason to prefer the sports page of a newspaper to its society page even though both pages have the same number of English words and the sports page has misprints, the society page none.

The partial ordering of information systems by their (gross) values will be studied. In particular, conditions, sufficient or necessary, will be stated under which two systems are comparable, so that one of them is "more in-

* Manuscript received February 28, 1966.

[1] This work was supported partly by the office of Naval Research under Task 047-041 and partly by the Western Management Science Institute under a grant from the Ford Foundation. Reproduction in whole or in part is permitted for any purpose of the United States Government. During 1964/5, Koichi Miyasawa, Professor of Statistics at the University of Tokyo, was post-doctoral Fellow under a Ford Foundation Grant to the University of California.

The article is the product of very close collaboration although a few of the sections are predominantly the results of the work of one of the two authors (e.g., Sections 1-7 in the case of Marschak [10, (12-4)]; and Theorem 8.2 in the case of Miyasawa). We owe thanks to Dr. Leif Appelgren, Stockholm, for the constructive criticism and to Professor Arthur Geoffrion of the Western Management Science Institute for many corrections and improvements.

[2] This cost will be assumed additive. We shall thus treat a special case only: see Section 4.

[3] See footnote 7 on the textbook terminology regarding "noiselessness" and "equivocation."

formative" than the other in the following sense: one of them can never have smaller value than the other for any payoff function defined on a given set of events. The ordering of information systems according to their informativeness has applications in the economics of information and organization; and also, as shown by D. Blackwell, in statistics (where messages and events correspond, respectively, to observations and hypotheses).[4]

Sections 1–5 will define concepts useful for the statement of our problem of finding conditions for the comparability of information systems. The remaining sections seek to solve this problem and to prove inclusion relations between the various conditions studied.

2. EVENTS AND DECISIONS

For a given actor (decision-maker), we define a set X of *states* x of the environment (not controlled by the actor[5]), a set C of *Consequence* c. Each function from X to C is called an act by Savage [**16**]. However, not all acts thus defined are feasible. Define the set A consisting of all feasible acts, a. These we shall call actions.[6] The set A can be thought of as the resources or technology available to the particular actor. Each action a maps X into C

$$a(x) = c\ .$$

This is, in general, a many-to-one mapping, for each state x may describe the environment in unlimited detail, and some of this detail may be irrelevant in the following sense. Two distinct states, x and x' may be such that $a(x) = a(x')$ for every a in A. We shall then say that x and x' are equivalent with respect to A. We can partition X into equivalence sets of the form

$$z_x^A \equiv \{x' \in X \colon a(x') = a(x),\ \text{all } a \in A\}\ . \tag{2.1.A}$$

Denote this partition by Z^A. Each equivalence set z^A in Z^A may be called an event, relevant with respect to technology A; or, briefly, an A-relevant event.

Example. Suppose X is an n-tuple of variables, some of them continuous, with n arbitrarily large; the consequences of any action of yours are not affected by many of these variables (political situation in Uganda) or by minor variations of the remaining ones (seconds of daily sunshine).

The set of states of the environment can be "coarsened" still further if we pay attention, not only to the actor's technology but also to his tastes. It would suffice for this purpose, to replace in (2.1.A) the equality sign between $a(x')$ and $a(x)$ by some symbol representing the actor's indifference between

[4] Blackwell [**2**], Blackwell and Girschick [**4**], also McGuire [**13**].

[5] Thus, in the case of sequential decisions, a state x would describe the time-seqence of external conditions. The sequence of states in the language of systems theory and dynamic programming corresponds (with the exception of the initial state) to a consequence in our terminology.

[6] In an earlier paper (Marschak [**11**]) the existence and uniqueness of the sets of payoff-relevant actions and events was proved, using a set of physically distinct feasible actions. We take this opportunity for a *correction* (due to K. Miyasawa): Theorem 3 of [**11**] is not valid, nor is its validity necessary for the proof of the main existence and uniqueness theorem.

these two consequences. There is no need to use numerical utilities. But since the numerical utility concept will be needed soon anyway, denote by u the actor's utility function, a function from C to the set of the reals, and write

$$(2.2) \qquad \omega(x, a) \equiv u(a(x)) \equiv u(c) ;$$

the function ω from $X \times A$ to the reals is called the *payoff function*. New partition X into equivalence sets of the form

$$(2.1.\omega) \qquad z_x^{\omega} \equiv \{x' \in X: \omega(x', a) = \omega(x, a), \text{ all } a \in A\} .$$

This partition, denoted by Z^{ω}, is clearly coarser than Z^A (i.e., Z^A is a sub-partition of Z^{ω}). We shall call z^{ω}, a typical equivalence set in Z^{ω}, an *event, relevant with respect to the payoff function* ω; or briefly, an ω-relevant event.

Example. Suppose each consequence of any of your actions is a triple (c_1, c_2, c_3) where c_1 and c_2 are your profits of this and of the next year, measured in cents; and c_3 is the amount of air pollution created by your plant. Suppose you are indifferent between $c \equiv (c_1, c_2, c_3)$ and $c' \equiv (c_1', c_2', c_3')$ because you are not concerned with differences less than \$1, or because you are willing to trade off a part of this year's profit for a profit increase in the next year, or because air pollution results only in other people's discomfort. Then $u(c) = u(c')$; and if, for all actions a in A, $a(x) = c$ and $a(x') = c'$, then the states x and x' belong to the same payoff-relevant event z^{ω}.

Considerations of "taste" induce also a coarsening of the set of actions. It may happen that, for two distinct actions a and a', and for all x in X,

$$u(a(x)) = u(a'(x)); \text{ that is, } \omega(x, a) = \omega(x, a') .$$

Let us, then, define a partition D^{ω} of A into equivalence sets of the form

$$(2.3) \qquad d_a^{\omega} \equiv \{a' \in A: \omega(x, a') = \omega(x, a), \text{ all } x \in X\} .$$

For convenience we shall call D^{ω} the set of *ω-relevant decisions*; its typical element d^{ω} will sometimes be denoted briefly as d.

Example. Let each consequence be a triple (c_1, c_2, c_3) defined in the previous example; suppose you are indifferent to air pollution and that two methods of production, a and a', always yield (i.e., for every x) the same profits but different amounts of air pollution. Then a and a' belong to the same ω-relevant decision d^{ω}.

Observing that ω is constant on $z^{\omega} \times d^{\omega}$, we may write without ambiguity $\omega(z^{\omega}, d^{\omega})$ where ω is defined over the domain $Z^{\omega} \times D^{\omega}$ instead of $X \times A$.

In what follows we shall be interested in varying the payoff function ω subject to a constraint depending on an arbitrarily fixed partition Z of X into *events* z. Given a payoff function ω, Z is called *ω-adequate* if it is a sub-partition of the *ω-relevant* partition Z^{ω}, see Marschak-Radner [**12**, (chapter 2)]; for example, Z^A (relevant with respect to technology but not necessary with respect to tastes) is ω-adequate. Then every $z \in Z$ is a subset of exactly one $z^{\omega} \in Z^{\omega}$ and we can write, without danger of ambiguity, $\omega(z, d^{\omega})$, where the function ω is defined over the domain $Z \times D^{\omega}$. Now let Ω_Z be the set of all payoff functions ω for which Z is ω-adequate. Given a fixed set Z of events we shall vary the payoff function ω over the set Ω_Z.

It has been shown—see Savage [16]—that certain plausible, quasi-logical postulates imply for a consistent decision-maker: (a) the existence of a numerical function u on the set C of consequences; hence a numerical function ω on $Z \times D^{\omega}$; and (b) the existence of a probability measure $\mathcal{P}$ on X; hence of a probability function $p(z)$ on Z, with the following property: given two decisions d and d' in D^{ω}, a consistent actor will not prefer d' to d if (assuming Z finite for simplicity)

$$\sum_{z \in Z} \omega(z, d)p(z) \geqq \sum_{z \in Z} \omega(z, d')p(z) . \tag{2.4}$$

The two compared averages are called the (expected) utilities of the decisions d and d', respectively. The proposition just stated is called the expected utility theorem. Roughly, it says that an actor conforming with certain consistency postulates (and with the rules of ordinary logic) maximizes the expected utility of decision.

We stated above that the set $A = \{a\}$ of actions (i.e., feasible acts) represents the actor's technology, and that the utility function u on the set C represents his tastes; hence the payoff function $\omega(x, a) \equiv u(a(x))$ reflects both. We can now add his beliefs, represented by the probability measure $\mathcal{P}$ on X. In what follows we shall consider $\mathcal{P}$ and Z fixed; and we shall permit ω to vary over the set Ω_Z for which Z is ω-adequate. This will enable us to discuss the "informativeness" of information systems for an arbitrary set of their users: see *Example* in Section 5.

3. INFORMATION SYSTEMS

An *information system* Y is a set consisting of (potential) *messages* y. We shall regard Y as another partition of the set X of states x. Unlike partition Z of X into payoff-adequate events z, the partition Y of X into messages y is not associated with the feasibility of actions and the indifference among their results. Instead. Y is associated with some object—an instrument—that produces messages. See Figures 1a–1c. *Example*: The state x may be such that (a) my barometer shows low pressure: this message y; and that (b) the visibility at the airport is low, thus affecting the success of a decision to fly: this is event z.

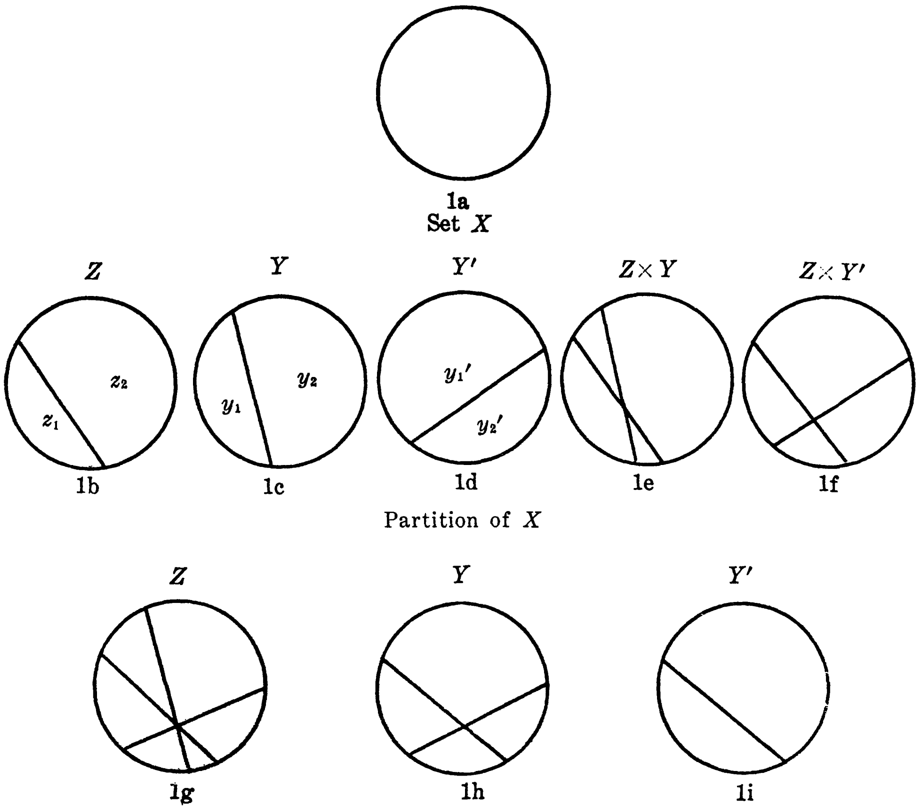
In the language of information theory, a set Z of events would be called "source" and a set Y of messages, a "channel." In the language of statistical inference, Z represents a set of alternative hypotheses, and Y represents the set of outcomes of an experiment and is itself called an experiment.

If a probability measure $\mathcal{P}$ is defined on X, the joint probability function on $Z \times Y$ is determined. In fact, given a set $(Y, Y', Y'', \cdots)$ of *available information systems*, the multivariate distribution on $Z \times Y \times Y' \times Y'' \times \cdots$ is defined. We shall write, using the same symbol $p(\ )$ for probability functions over different domains, yet without risk of ambiguity:

probability of event,

$$\Pr(x \in z) \equiv p(z) ;$$

probability of message,



Noiseless Cases ($Z \,\mathrm{s}\, Y$, $Z \,\mathrm{s}\, Y'$), with $Y \,\mathrm{s}\, Y'$

FIGURE 1

SET X AND ITS PARTITIONS

$$\Pr(x \in y) \equiv p(y)\ ,$$

($p(z)$ and $p(y)$ are all positive since z and y are non-empty);

joint probability of event and message,

$$\Pr(x \in z \cap y) \equiv p(z, y)\ ;$$

posterior probability of event, given the message,

$$p(z, y)/p(y) \equiv p(z \mid y)\ ;$$

likelihood of message, given the event,

$$p(z, y)/p(z) \equiv p(y \mid z)\ .$$

When comparing two information systems Y and Y' we shall also use expressions like the following:

probability of joint occurrence of message $y \in Y$ and $y' \in Y'$,

$$\Pr(x \in y \cap y') = p(y, y')\ ;$$

probability of joint occurrence of event z and messages y, y',

$$\Pr(x \in z \cap y \cap y') = p(z, y, y')\ ;$$

probability of y', given y,

$$p(y, y')/p(y) \equiv p(y' \mid y)\ ;$$

posterior probability of event z, given messages y, y',

$$p(z, y, y')/p(y, y') \equiv p(z \mid y, y')\ ;$$

and so on.

For simplicity of reasoning, we have assumed the set Z of events and the sets $Y, Y', \cdots$, of messages *finite*. (No such assumption will be made about X, the set of states, except in Theorem 11.4.) Specifically,

$$Z \equiv (z_1, \cdots, z_m); \quad Y \equiv (y_1, \cdots, y_n); \quad Y' \equiv (y'_1, \cdots, y'_{n'}); \quad m, n, n', \cdots \geqq 2\ .$$

The generic elements z of Z, y of Y, y' of Y', $\cdots$, can be regarded as random variables taking, respectively, the values

$$z_i (i = 1, \cdots, m); \quad y_j (j = 1, \cdots, n); \quad y'_k (k = 1, \cdots, n'); \cdots\ .$$

In most of our discussion Z will be fixed, and the effects of varying Y (i.e., of replacing it by Y', say) will be studied—see Figures 1b–1f—using for the three marginal probabilities the following vectors (the various alternative notations will be used according to convenience):

$$\begin{aligned}
&\text{the } (m \times 1)\text{-vector } [p(z)] \equiv [p(z_i)] \equiv [r_z] \equiv [r_i] \equiv r \\
&\text{the } (n \times 1)\text{-vector } [p(y)] \equiv [p(y_j)] \equiv [q_j] \equiv q^Y \equiv q \\
&\text{the } (n' \times 1)\text{-vector } [p(y')] \equiv [p(y'_k] \equiv [q'_k] \equiv q^{Y'} \equiv q'\ .
\end{aligned}$$

For the joint probabilities of events and messages; the posterior probabilities of events (given the message in Y); and the likelihoods of those messages (given the events in Z), we use the $(m \times n)$-matrices

$$\begin{aligned}
[p(z, y)] &\equiv [p(z_i, y_j)] \equiv [p_{zy}] \equiv [p_{ij}] \equiv P^Y \equiv P \\
[p(z \mid y)] &\equiv [p(z_i \mid y_j)] \equiv [\pi_{zy}] \equiv [\pi_{ij}] \equiv \Pi^Y \equiv \Pi \\
[p(y \mid z)] &\equiv [p(y_j \mid z_i)] \equiv [\lambda_{zy}] \equiv [\lambda_{ij}] \equiv \Lambda^Y \equiv \Lambda
\end{aligned}$$

and the corresponding $(m \times n')$-matrices for Y'

$$P^{Y'} \equiv P'; \quad \Pi^{Y'} \equiv \Pi'; \quad \Lambda^{Y'} \equiv \Lambda'\ .$$

By definition (writing henceforth $\sum_y, \sum_z$ for summation over sets Y, Z)

$$\text{(3.1.1)} \qquad \sum_z p_{zy} = q_y > 0, \quad \sum_y p_{zy} = r_z > 0, \quad \pi_{zy} \geqq 0, \quad \lambda_{zy} \geqq 0$$

$$\text{(3.1.2)} \qquad p_{zy} = \lambda_{zy} r_z = \pi_{zy} q_y \geqq 0$$

$$\text{(3.1.3)} \qquad 1 = \sum_y q_y = \sum_z r_z = \sum_y \lambda_{zy} = \sum_z \pi_{zy} = \sum_y \sum_z p_{zy}\ .$$

Moreover, with Z fixed, r is fixed and we have for any Y, Y'

$$\text{(3.2)} \qquad \Pi q = r = \Pi' q'\ .$$

We shall also use the $(n \times n')$-matrix

$$(3.3) \qquad [p(y' \mid y)] \equiv [p(y'_k \mid y_j)] \equiv [\gamma_{yy'}] \equiv [\gamma_{jk}] \equiv \Gamma \, .$$

Clearly,

$$(3.4) \qquad \gamma_{yy'} \geqq 0; \quad \sum_{y'} \gamma_{yy'} = 1 \, ;$$

$$(3.5) \qquad q' = \Gamma_{(\tau)} q \, ,$$

where $\Gamma_{(\tau)}$ is the transpose of Γ. Similarly, we define the $(n' \times n)$-matrix

$$(3.3') \qquad [p(y \mid y')] \equiv [p(y_j \mid y'_k)] \equiv [\gamma'_{yy'}] \equiv [\gamma'_{jk}] \equiv \Gamma' \, ,$$

$$(3.4') \qquad \gamma'_{yy'} \geqq 0; \quad \sum \gamma'_{yy'} = 1 \, ,$$

$$(3.5') \qquad q = \Gamma' q' \, .$$

Let T and T' be two partitions of X. T is said to be a *subpartition* of (or *finer than*) T' (or, equivalently, T' is coarser than T),

$$(3.6) \qquad T \, \mathrm{s} \, T'$$

if each t in T is contained in one of the t' in T'; or equivalently, there is a many-to-one correspondence $T \to T'$. Condition (3.6) implies—and, in case of X finite, is equivalent to—the following: for any $t \in T, t' \in T'$,

$$(3.7) \qquad p(t' \mid t) = \begin{cases} 1 & \text{if } t \subset t' \\ 0 & \text{otherwise} \, . \end{cases}$$

In particular, if the set Z of events is finer than the information system Y, (see Figures 1g, 1h)

$$(3.8) \qquad Z \, \mathrm{s} \, Y \, ,$$

we shall say that Y is *noiseless* (with respect to Z). In this case there is a *many-to-one correspondence* $Z \to Y$; and it will follow that

$$(3.9) \qquad p(y \mid z) \equiv \lambda_{zy} = \begin{cases} 1 & \text{if } z \subset y \\ 0 & \text{otherwise} \, . \end{cases}$$

Thus each row of Λ^Y consists of one 1 and $n - 1$ zeros if Y is noiseless.[7]

In Section 11, we shall consider the case when the two compared informa-

[7] This agrees with the terminology of Feinstein [7, (23)]; but Abramson [1, (11-2)] calls the case (3.10) a "deterministic channel," and reserves the adjective "noiseless" for the case when the correspondence $Z \to Y$ is one-to-many, hence in our notation,

$$p(z \mid y) = \Pi_{zy} = \begin{cases} 1 & \text{if } y \subset z \\ 0 & \text{otherwise} \, , \end{cases}$$

so that each column of Π^Y consists of 1 one and $m-1$ zeros. In this case, however, it should be easy (i.e., costless) for the user to consider all messages corresponding to the same event, as one and the same message; this establishes a *one-to-one* correspondence between the set of messages (thus redefined) and the set of events. This is sometimes called perfect information. Then Π and Λ are the same (permutation) matrix:

$$\lambda_{zy} = \pi_{zy} = \begin{cases} 1 & \text{if } z = y \\ 0 & \text{otherwise} \, , \end{cases}$$

It seems to us therefore that the case called "noiseless" by Abramson is of little interest: it collapses into the case of perfect information which is a specical case of noiseless information, in Feinstein's and our sense.

tion systems, Y and Y' are both *noiseless* (see Figures 1g, 1h, 1i):
Condition (N): $Z \mathrm{s} Y, Z \mathrm{s} Y'$.

If two information systems Y and Y' (noiseless or not) are such that Y is finer than Y', and thus Y' *coarser* than Y we write
Condition (C): $Y \mathrm{s} Y'$.

In this case we can also say that Y' is obtained from Y by *collapsing*, or *condensing*, several messages in Y into a single message in Y'. Under condition (C) there is a many-to-one correspondence $Y \to Y'$, and (C) will imply, as in (3.7) that

$$p(y' \mid y) \equiv \gamma_{yy'} = \begin{cases} 1 & \text{if } y \subset y' \\ 0 & \text{otherwise}. \end{cases} \tag{3.10}$$

4. INFORMATION VALUES: GROSS AND NET

The actor associates each message in Y with some decision in D^{ω}. This mapping will be called a *decision* rule, δ^{ω}, an element of a set Δ^{ω}. Without danger of ambiguity we shall often omit the superscript ω. Thus

$$\begin{aligned} d &= \delta(y) , \\ \omega(z, d) &= \omega(z, \delta(y)); \qquad z \in Z, y \in Y, d \in D, \delta \in \Delta . \end{aligned} \tag{4.1}$$

Thus, given the payoff function ω, the utility of the result depends on the event z, the message y and the decision rule δ.

The utility amount $\omega(z, \delta(y))$ may be interpreted by expressing the economic effect (i.e., the effect upon the utility to the decision-maker), not only of decision $\delta(y)$, given event z, but also of the decision rule δ itself. For example, a simple (e.g., linear) decision rule is less costly to apply than a complicated one. Moreover, a decision rule from Y to D, where D consists of feasible decisions (actions), may itself be non-feasible—for example, if it is so complicated as to exceed the decision-making capabilities of available men or machines. Strictly speaking, we should define a feasible subset Δ_{φ} of Δ, for use in any further economic discussion of information systems. For the sake of simplicity we shall neglect in most of the following both the cost and the possible non-feasibility of decision rules and assume $\Delta_{\varphi} = \Delta$.

An optimal decision rule δ^* in Δ maximizes the utility averaged over all messages y in Y. The maximum expected utility thus achieved will be denoted by $U(P^Y; \omega)$ for it will depend on the joint probability P^Y of y and z (with Z fixed) and on the payoff function ω:

$$\begin{aligned} U(P^Y; \omega) &\equiv \sum_y \sum_z \omega(z, \delta^*(y)) p_{zy} \\ &\geqq \sum_y \sum_z \omega(z, \delta(y)) p_{zy} \end{aligned} \tag{4.2}$$

for all δ in Δ. Then, since $p_{zy} = q_y \pi_{zy}$, we have

$$\begin{aligned} U(P^Y; \omega) &\equiv \max_{\delta \in \Delta} \sum_y q_y \sum_z \pi_{zy} \omega(z, \delta(y)) \\ &\equiv \sum_y q_y \max_{d \in D} \sum_z \pi_{zy} \omega(z, d) \\ &\equiv \sum_y q_y \sum_z \pi_{zy} \omega(z, d_y) \equiv \sum_y \sum_z p_{zy} \omega(z, d_y) , \end{aligned} \tag{4.3}$$

thus defining $d_y \equiv \delta^*(y)$ as the optimal decision in response to message y:

$$\sum_z \pi_{zy}\omega(z, d_y) \geqq \sum_z \pi_{zy}\omega(z, d) , \qquad \text{all } d \in D . \tag{4.4}$$

In order to guarantee the achievement of a maximum in (4.3), (4.4), we shall henceforth assume that the set of real numbers $\omega(z, d), d \in D^\omega$, is bounded from above and closed and that ω is continuous.

A more appropriate and precise name for the quantity $U(P^Y; \omega)$ would be *gross* value of Y, in contrast to its *net* value. So far, we have neglected those economic effects of an information instrument that are due not to decision based on messages received through the instrument but to its other properties, roughly called "cost of information." For example, one of two compared instruments may be more expensive, or it may send messages that while related to events by the same probability distribution are more time-consuming. Strictly speaking, these two instruments produce two different partitions of X: for the occurrence of a particular message belongs to the detailed description of a state x; and the consequence may depend, not only on decision d and event z but also on the manner in which the message received through a particular instrument affects the user's resources and well-being, through the expenditure of time, money and effort involved in the process of receiving the message. Equation (3.2) should be replaced by a more general one

$$u(c) \equiv u(d(z), y) \equiv \omega_Y(z, d, y) \tag{4.5}$$

say, where y belongs to a set Y of messages associated with a particular instrument. The new function ω_Y may be called the *net payoff function*. In a simple but important case

$$\omega_Y(z, d, y) = \omega(z, d) - \kappa(y) , \tag{4.6}$$

where, for example, $\kappa(y)$ is the monetary cost associated with message y, $\omega(z, d)$ is the ("gross") monetary profit from decision d when event z occurs, and utility is linear in money: see Raiffa and Schlaifer [**15**, (section 4)]. In general the net value of an information system Y can be written as

$$U_Y(Y) \equiv \max_{\delta \in \Delta} \sum_z \sum_y \omega_Y(z, \delta(y), y)p_{zy} \tag{4.7}$$

so that in the special, additive case (4.6) we have,

$$U_Y(Y) = U(P^Y; \omega) - \sum_y \kappa(y)p(y) . \tag{4.8}$$

In what follows we shall indeed assume (4.6), hence (4.8). This will permit us to study the economic effect of statistical properties of an information system, separated from the system's cost properties. But remember that in the general case (4.5), (4.7), the net value of an information system cannot be decomposed into gross payoff and cost as additive components.

5. COMPARISON OF (GROSS) INFORMATION VALUES

We shall compare the (gross) values of two information systems, Y and Y', having fixed the set Z of events and thus the vector $r = [r_z]$ of marginal

probabilities of events and the set Ω_Z of considered payoff functions.

Example. To expand further the flight passenger's case of Section 2: We can ask whether a barometer or a hygrometer (with messages referring to humidity) is preferable to a user who has a ranch, an airplane, or both. The uses considered will define the set Z of payoff-adequate weather descriptions which will permit the necessary comparison.

In the ("Bayesian") spirit of the expected utility theorem, the two joint distributions P^Y and $P^{Y'}$ (also denoted as P and P') are known to the decision maker. He knows therefore also the pairs

$$(\Pi_*, q_*) \text{ and } (\Pi'_*, q'_*) ,$$
$$(\Lambda_*, r_*) \text{ and } (\Lambda'_*, r_*) ,$$

indicating by the star subscript (as it will be sometimes useful to do) that those matrices and vectors are known. Such knowledge is usually assumed in operations research problems—including the most general ones of sequential decision making, in which one starts with an initial distribution of states seen as time sequences and conditional upon the sequence of decisions: see Miyasawa [14]. The knowledge of r_* is also often assumed in information theory: for each channel Y, the joint distribution $p(z, y)$ is known, permitting one to compute, for a given source Z, the equivocation characterizing the channel Y[8]

$$H(Z \mid Y) = -\sum_y p(y) \sum_z p(z \mid y) \log p(z \mid y) . \tag{5.1}$$

(Compare this with (4.3)!)

Also, $U(P^Y; \omega)$ and $U(P^{Y'}; \omega)$ can sometimes be ordered, for all ω considered without the knowledge of the distributions P^Y aed $P^{Y'}$ and only knowing the process that produces Y and Y' and imposes certain general restrictions on the trivariate distribution of z, y, y'. Such will be, in fact, the content of Sections 6 and 7.

It is also useful to consider where the comparison between information values of Y and Y' is made by someone who uses the knowledge of $\Pi = \Pi_*$ and $\Pi' = \Pi'_*$ (ignoring or being ignorant of q, q'); or who uses the knowledge of $\Lambda = \Lambda_*$ and $\Lambda' = \Lambda'_*$ (ignoring or being ignorant of the value of r). Such cases of limited knowledge will be considered because they may simplify the procedure of comparing information values. In addition, an actor may have knowledge of posterior probabilities of π_{zy} associated with each message y, but have trouble ascertaining the probability q_y of each message: this makes the comparison of Π with Π' interesting. (Note, however, that the probability vectors q, q', r even though ignored or not known are constrained by identity (3.2).)

[8] The equivocation $H(Z \mid Y) = 0$ if and only if all $p(z \mid y) \equiv \pi_{zy} = 1$ or 0, i.e., if and only if Y is noiseless in the sense of Abramson and not in the sense used by Feinstein and ourselves. On the other hand, Y is noiseless in the latter sense, i.e., all $p(y \mid z) \equiv \lambda_{zy} = 1$ or 0, if and only if the expression

$$H(Y \mid Z) \equiv -\sum_z p(z) \sum_y p(y \mid z) \log p(y \mid z) = 0 .$$

Following Abramson, the latter expression might be called equivocation of Y with respect to Z.

Finally the ignorance of the (prior) distribution of hypotheses is maintained in non-Bayesian statistics, making it impossible to use other than the likelihood matrices $\Lambda = \Lambda_*$ and $\Lambda' = \Lambda'_*$.

Whenever $P^Y, P^{Y'}$ are known (hence $r, \Lambda, \Lambda', \Pi, \Pi'$ have known values r_*, Λ_*, etc.) and are such that

$$(5.2) \qquad U(P^Y; \omega) \geqq U(P^{Y'}; \omega), \qquad \text{all } \omega \text{ in } \Omega_Z,$$

we shall write

Condition (P): $P^Y > P^{Y'}$ (or briefly: $P > P'$),

and say: "Y is more informative[9] than Y' as revealed by their joint distributions P, P' with Z" or simply "*P is more informative than P'.*" Note that (5.2) can be rewritten thus

$$(5.2.1) \qquad U(\Pi_*, q_*; \omega) \geqq U(\Pi'_*, q'_*; \omega), \text{ all } \omega \text{ in } \Omega_Z; \text{ or } (\Pi_*, q_*) > (\Pi'_*, q'_*),$$

or

$$(5.2.2) \qquad U(\Lambda_*, r_*; \omega) \geqq U(\Lambda'_*, r_*; \omega), \text{ all } \omega \text{ in } \Omega_Z; \text{ or } (\Lambda_*, r_*) > (\Lambda'_*, r_*).$$

If r is ignored or not known, we may ask whether

$$(5.3) \qquad U(\Lambda_*, r; \omega) \geqq U(\Lambda'_*, r; \omega), \qquad \text{all } \omega \text{ in } \Omega_Z \text{ and all } r;$$

and, if so, write (omitting the stars for brevity)

Condition (Λ): $\Lambda > \Lambda'$

and say: "Y is more informative than Y' as revealed by the likelihood matrices Λ and Λ'," or simply: "*Λ is more informative than Λ'.*"

Finally, it may also be of interest to ask whether or not

$$(5.4) \qquad U(\Pi_*, q; \omega) \geqq U(\Pi'_*, q'; \omega) \text{ for all } \omega \text{ in } \Omega_Z \text{ and all } q, q' \text{ such that } \Pi_* q = \Pi'_* q'.$$

This question arises naturally when the marginal probabilities q, q', r are ignored or not known. Since Z, hence r, is the same for both information systems, $\Pi_* q = \Pi'_* q'$ by (3.2). When (5.4) is satisfied we shall say (omitting stars for brevity) that "Y is more informative than Y' as revealed by the posterior probabilities Π, Π'"; or briefly "*Π is more informative than Π',*" and write

Condition (Π): $\Pi > \Pi'$.

The binary relation "$>$" defined above on the set of all matrices P, or all matrices Π, or all matrices Λ, as the case may be, induces in all cases at most a partial ordering on the P (or the Π or the Λ), hence on any set of available information systems. For, in general, there will be some pairs (P, P') or (Π, Π') or (Λ, Λ') for which there exist two payoff functions, ω and ω' in Ω_Z, such that the inequality in (5.2) or (5.3) or (5.4) is valid for ω but not for ω'. In contrast, the equivocation (5.1) used in information theory does not depend on the payoff function and thus induces complete (and weak)

[9] Strictly speaking, the term "not less informative" would be more appropriate. We have chosen terms that will be shown to harmonize with those of Blackwell: see Section 13.

ordering. It will be shown that lower equivocation is necessary but not sufficient for higher informativeness: see Section 12.

Since q, q' and r are all fixed (at starred values) in condition (P) but not in condition (Λ) nor in condition (Π), condition (P) is weaker than either of the other two, and we have

THEOREM 5.1. $(\Lambda) \Rightarrow (P); (\Pi) \Rightarrow (P)$.

We shall now prove that (Λ) and (P) are equivalent.

THEOREM 5.2. $(P) \Leftrightarrow (\Lambda)$.

PROOF. By Theorem 5.1, it will be enough to prove $(P) \Rightarrow (\Lambda)$. Let $P^Y \equiv (\Lambda, r)$, $P^{Y'} \equiv (\Lambda', r)$ and assume that, as in (5.2.2) but with stars omitted for brevity, condition (P) holds. From (3.1.2) and (4.3), for any payoff function $\omega \in \Omega_Z$ from $D^\omega \times Z$ to the reals,

$$
\begin{aligned}
U(P^Y; \omega) &\equiv U(\Lambda, r; \omega) \\
&\equiv \sum_y \max_{d \in D^\omega} \sum_z r_z \lambda_{zy} \omega(z, d) \,.
\end{aligned}
\tag{5.5}
$$

For any given $r = \bar{r}$ define a new payoff function $\bar{\omega}$ in Ω_Z by

$$
\bar{\omega}(z, d) = \frac{\bar{r}_z}{r_z} \omega(z, d); \; D^{\bar{\omega}} = D^\omega \,.
\tag{5.6}
$$

Then, replacing ω by $\bar{\omega}$ in (5.5),

$$
\begin{aligned}
U(P^Y; \bar{\omega}) &= \sum_y \max_{d \in D^\omega} \sum_z \bar{r}_z \lambda_{zy} \omega(z, d) \\
&= U(\Lambda, \bar{r}; \omega) \,.
\end{aligned}
\tag{5.7}
$$

By similar reasoning,

$$
U(P^{Y'}; \bar{\omega}) = U(\Lambda', \bar{r}; \omega) \,.
\tag{5.7$'$}
$$

From condition (P), we know

$$
U(P^Y; \bar{\omega}) \geqq U(P^{Y'}; \bar{\omega}) \,.
\tag{5.8}
$$

Therefore by (5.7), (5.7$'$) and (5.8),

$$
U(\Lambda, \bar{r}; \omega) \geqq U(\Lambda', \bar{r}; \omega) \,,
\tag{5.9}
$$

where $\bar{r}$ and ω are arbitrary, so that (5.3) is satisfied. Thus (P) implies (Λ).

It follows from Theorem 5.2 that, if one system is more informative than another as revealed by their respective likelihood matrices, no further knowledge is added by knowing the prior probabilities of events. On the other hand, we shall show in Section 9, Theorem 9.3 that (P) implies (Π) only under certain conditions. Anticipating this result we state a summarizing

THEOREM 5.3. $(\Pi) \mapsto (P) \Leftrightarrow (\Lambda)$,

using here and henceforth the symbol $\mapsto$ for "implies but is not implied by."

We conclude this section by remarking that it is not possible to apply the relation "more informative than" to net rather than gross information values. Modifying the notation of (4.7) in an obvious way, suppose that the net payoff functions ω_Y and $\omega_{Y'}$ defined respectively on $Z \times D \times Y$ and $Z \times D \times Y'$ are such that for some positive number k

(5.10) $$k > U_Y(Y; \omega_Y) - U_{Y'}(Y'; \omega_{Y'}) \geqq 0\,;$$

then there exists another pair of net payoff functions, $\bar{\omega}_Y$ and $\bar{\omega}_{Y'}$, such that

$$U_Y(Y; \bar{\omega}_Y) - U_{Y'}(Y'; \omega_{Y'}) < 0:$$

for example, let

$$\bar{\omega}_Y(z, d, y) = \omega_Y(z, d, y) - \frac{1}{2}k\,; \qquad \bar{\omega}_Y\,(z, d, y') = \omega_Y\,(z, d, y') + \frac{1}{2}k\,.$$

In particular, when the net payoffs can be decomposed into gross payoffs and information costs as in (4.6), it is always possible to imagine information cost functions $\kappa(y), \kappa'(y')$ that would reverse the second inequality in (5.10).

6. GARBLED INFORMATION

For any joint distribution of three variables z, y, y' the following identities hold:

(6.1) $$\frac{p(y' \mid y, z)}{p(y' \mid y)} \equiv \frac{p(z \mid y, y')}{p(z \mid y)} \equiv \frac{p(y, y' \mid z)}{p(y \mid z) \cdot p(y' \mid y)}\,,$$

since each of these three ratios is equal to

$$\frac{p(z, y, y') \cdot p(y)}{p(z, y) \cdot p(y, y')}\,.$$

We shall say that, for a given set Z of events, the information system Y in *garbled* into Y' when, for all $z \in Z, y \in Y, y' \in Y$ the following condition holds:

Condition (G): *Each of the three ratios in* (6.1) *is equal to* 1.

(G) can therefore be written in three equivalent forms. Each seems to agree with the ordinary usage of the term garbling.[10]

(G_1) $$p(y' \mid y, z) = p(y' \mid y)\,,$$

i.e., given message y of the original information system Y, the conditional probability of message y' of the garbled system Y' does not depend on event z.

(G_2) $$p(z \mid y, y') = p(z \mid y)\,,$$

i.e., the posterior probability of event z given the original message y, does not depend on the garbled message y'.

(G_3) $$p(y, y' \mid z) = p(y \mid z) \cdot p(y' \mid y):$$

this describes the generation of the pair of messages, y and y': while in general $p(y, y' \mid z) = p(y \mid z) \cdot p(y' \mid y, z)$, this identity becomes in the garbling case, (G_3), because of (G_1).

The following condition that is implied, but does not imply[11] condition (G), is obtained by summing (G_3) over y

[10] This is called "cascade" in information theory: e.g., Abramson [1, (section 5.9)].

[11] A rigorous proof will be given in Section 9.

$$p(y' \mid z) \equiv \sum_y p(y, y' \mid z) = \sum_y p(y \mid z) p(y' \mid y) , \qquad \text{all } z \in Z, y' \in Y' ;$$

or, in the notation of Section 2,

$$\lambda'_{zy'} = \sum_y \lambda_{zy} \gamma_{yy'} . \tag{6.2}$$

Using matrix notation we rewrite this as

Condition (Γ): $\Lambda' = \Lambda\Gamma$,
and obtain

THEOREM 6.1. $(G) \mapsto (\Gamma)$.

A row-stochastic, or Markov, matrix is one with non-negative elements only, and with all row-sums=1. Denote the class of all Markov matrices of order $n \times n'$ by $\mathscr{M}$, and the class of their transposes, i.e., of all "column-stochastic" matrices of order $n' \times n$, by $\mathscr{M}_{(\tau)}$. Then obviously

$$\Gamma \in \mathscr{M}, \Gamma' \in \mathscr{M}_{(\tau)} . \tag{6.3}$$

In honor of David Blackwell, we have called the following condition

Condition (B): There exists a matrix $B = [\beta_{jk}]$ such that

$$\begin{array}{ll} (\mathrm{B}_0) & B \in \mathscr{M} \\ (\mathrm{B}_1) & \Lambda' = \Lambda B . \end{array} \tag{6.4}$$

Condition (B) is implied by the garbling condition (G), but not conversely; for we can prove

THEOREM 6.2. $(G) \mapsto (\Gamma) \mapsto (B)$.

PROOF. By (6.3), (Γ) implies (B). The converse is not necessarily true (see Section 9). Hence our theorem follows from Theorem 6.1. We shall now prove the important

THEOREM 6.3. $(B) \Rightarrow (\Lambda)$.

PROOF. For any $(\Lambda, r) \equiv (\Pi, q)$ and ω, define for every message y_j the optimal decision d_j; that is, by (4.4),

$$\sum_{i=1}^{m} \pi_{ij} \omega(z_i, d_j) \geqq \sum_{i=1}^{m} \pi_{ij} \omega(z_i, d) , \qquad \text{all } d \in D^\omega; \tag{6.5}$$

or, omitting the summation limits for brevity and writing

$$\omega(z_i, d_j) \equiv u_{ij} ,$$

$$\sum_i \pi_{ij} u_{ij} \geqq \sum_i \pi_{ij} \omega(z_i, d) , \qquad \text{all } d \in D^\omega , \tag{6.6}$$

$$\sum_i r_i \lambda_{ij} u_{ij} \geqq \sum_i r_i \lambda_{ij} \omega(z_i, d) , \qquad \text{all } d \in D^\omega , \tag{6.7}$$

by (3.1.2). Then by (4.3)

$$\begin{aligned} U(P; \omega) &\equiv U(\Lambda, r; \omega) \equiv U(\Pi, q; \omega) \\ &= \sum_{i,j} q_j \pi_{ij} u_{ij} = \sum_{i,j} r_i \lambda_{ij} u_{ij} . \end{aligned} \tag{6.8}$$

Similarly, for $(\Lambda', r) \equiv (\Pi', q')$, we define $d'_k \in D^\omega$ and $u'_{ik} \equiv \omega(z_i, d_k)$ so that

(6.6′) $$\sum_i \pi'_{ik}u'_{ik} \geqq \sum_i \pi'_{ik}\omega(z_i, d)\,, \qquad \text{all } d \in D^{\omega}\,,$$

(6.7′) $$\sum_i r_i\lambda'_{ik}u'_{ik} \geqq \sum_i r_i\lambda'_{ik}\omega(z_i, d)\,, \qquad \text{all } d \in D^{\omega}\,,$$

and we have

(6.8′) $$\begin{aligned} U(P'; \omega) &\equiv U(\Lambda', r; \omega) \equiv U(\Pi', q'; \omega) \\ &= \sum_{i,k} q'_k\pi'_{ik}u'_{ik} = \sum_{i,k} r_i\lambda'_{ik}u'_{ik}\,. \end{aligned}$$

Suppose (B) holds. Then by (6.8′), (6.7), (6.8) we have, for any r, ω,

$$\begin{aligned} U(\Lambda; r; \omega) &= \sum_{j,k} \beta_{jk} \sum_i r_i\lambda_{ij}u'_{ik} \\ &\leqq \sum_{j,k} \beta_{jk} \sum_i r_i\lambda_{ij}u_{ij} = \sum_{i,j} r_i\lambda_{ij}u_{ij} \\ &= U(\Lambda, r; \omega)\,. \end{aligned}$$

This proves that (B) implies (Λ). In Section 8 we shall prove the converse, so that (B) $\Leftrightarrow$ (Λ).

From Theorems 6.1, 6.2 and 6.3 we immediately obtain

THEOREM 6.4. *If Y is garbled into Y' then Y is more informative than Y' as revealed by their likelihood matrices Λ and Λ'.*

The following theorem summarizes the inclusion relations stated in this and the preceding section:

THEOREM 6.5. $(G) \mapsto (\Gamma) \mapsto (B) \Leftrightarrow (\Lambda) \Leftrightarrow (P) \Leftarrow\!\!\dashv (\Pi)$.

The proof of this theorem will be complete when, as mentioned before, it is shown that (Λ) implies (B) (Section 8) and that (B) does not imply (Γ) nor does (Γ) imply (G) (Section 9). Pending these proofs we have proved so far only the weaker

THEOREM 6.6. $(G) \Rightarrow (\Gamma) \Rightarrow (B) \Rightarrow (\Lambda) \Leftrightarrow (P)$.

7. COLLAPSING INFORMATION AND JOINING INFORMATION

The case in which Y' is condensed from Y was defined at the end of Section 3 as

(C) $$Y \mathrm{s} Y'\,,$$

i.e., each y in Y is contained in one y' in Y'. Hence by (2.7)

$$\begin{aligned} &\text{either} \quad y \subset y', \;\; p(y' \mid y) = 1, \;\; p(y, y' \mid z) = p(y \mid z) \\ &\text{or} \qquad y \not\subset y', \;\; p(y' \mid y) = 0, \;\; p(y, y' \mid z) = 0\,. \end{aligned}$$

In either case, (G_1) is satisfied. Hence (C) implies (G); but the converse is obviously not true. Collapsing information is thus a special case of garbling. We state this as

THEOREM 7.1. $(C) \mapsto (G)$.

Let every message in Y be obtained by *joining* a message in Y' with a message belonging to a third information system, T. Then,[12] since Y, Y', T

[12] See, e.g., Halmos [8].

are partitions of X,

Condition (J): $Y = Y' \times T = \{(y' \cap t)\},\ y' \in Y', t \in T$.
Clearly Y is finer than Y', so that (J) implies (C); but the converse is, of course, not true, and we have

THEOREM 7.2. $(J) \Rightarrow (C)$.
We can combine these results with those of Sections 5 and 6 into

THEOREM 7.3. $(J) \Rightarrow (C) \Rightarrow (G) \Rightarrow (\Gamma) \Rightarrow (B) \Rightarrow (\Lambda) \Leftrightarrow (P)$.
Thus Y' is less informative than Y when Y' is garbled or condensed from Y; or when Y is formed by joining Y' with a message from a third information system. Garbling and condensing information occurs when an intermediate node is inserted in a communication network—as, for example, when reports are processed into commands. Joining information occurs when information is centralized. It follows that inserting an intermediate node never increases, and centralizing information never decreases, the gross expected payoff. But the *net* payoffs may be ordered differently. For example, the cost of centralizing information may offset its advantages (see remarks at the end of Section 5) or such centralization may call for rules that are not feasible (cf. Section 4).

8. EQUIVALENCE OF CONDITIONS (B), (Λ), (P) AND (θ)

THEOREM 8.1. $(B) \Leftrightarrow (P) \Leftrightarrow (\Lambda)$.

PROOF. Because of Theorem 5.2 and 6.3, it will suffice to prove $(P) \Rightarrow (B)$. We shall first prove the following:

LEMMA. If (P) holds, then for any real $(n' \times m)$-matrix $V = [v_{ki}]$ there exists a matrix $M = [m_{kj}] \in \mathscr{M}_{(\tau)}$ such that

$$(8.1) \qquad \sum_{i,k,j} m_{kj} r_i \lambda_{ij} v_{ik} \geqq \sum_{i,k} r_i \lambda'_{ki} v_{ki}\ .$$

To prove the lemma let us define, for a given matrix $V = [v_{ki}]$ the payoff function ω as follows: $D^\omega = \{d_1, \cdots, d_{n'}\}$, and

$$(8.2) \qquad \omega(z_i, d_k) = v_{ki}\ , \qquad i = 1, \cdots, m;\ k = 1, \cdots, n'\ .$$

Now define, for each message y_j in Y, the optimal decision $d_{k(j)} \in D^\omega$ by

$$(8.3) \qquad \sum_i r_i \lambda_{ij} \omega(z_i, d_{k(j)}) \geqq \sum_i r_i \lambda_{ij} v_{ki}\ , \qquad k = 1, \cdots, n';\ j = 1, \cdots, n\ ;$$

then by (4.3), (2.1.2)

$$(8.4) \qquad U(P^Y; \omega) = \sum_{i,j} r_i \lambda_{ij} v_{k(j)i}\ .$$

Now, for any matrix $T = [t_{kj}] \in \mathscr{M}_{(\tau)}$, we have by (8.3)

$$(8.5) \qquad \sum_{i,j,k} t_{kj} r_i \lambda_{ij} v_{k(j)i} = \sum_{i,j} r_i \lambda_{ij} v_{k(j)i} \geqq \sum_{i,j,k} t_{kj} r_i \lambda_{ij} v_{ki}\ ,$$

since $t_{kj} \geqq 0$, $\sum_k t_{kj} = 1$. That is

(8.6) $$U(P^Y; \omega) \geqq \sum_{i,j,k} t_{kj} r_i \lambda_{ij} v_{ki} .$$

Now by definition,

(8.7) $$U(P^{Y'}; \omega) \geqq \sum_{i,k} r_i \lambda'_{ik} \omega(z_i, d_k) , \qquad \text{all } d_k \in D^\omega, \text{ i.e.,}$$

(8.8) $$U(P^{Y'}; \omega) \geqq \sum_{i,k} r_i \lambda'_{ik} v_{ki} .$$

If condition (P) holds then,

(8.9) $$U(P^Y; \omega) \geqq U(P^{Y'}; \omega) .$$

Let our matrix T be such that

$$t_{kj} = \begin{cases} 1 & \text{if } k = k(j) \\ 0 & \text{otherwise.} \end{cases}$$

Then (8.6) becomes an equality. Therefore by (8.8) and (8.9) T has the properties of the matrix M required in the lemma.

To complete the proof of the theorem, let $\mathscr{V}$ be the set of all $(n' \times m)$-matrices $V = [v_{ki}]$ such that $0 \leqq v_{ki} \leqq 1$; as before, let $M = [m_{kj}]$ belong to $\mathscr{M}_{(\tau)}$; and define a function $\Psi(M, V)$ on $\mathscr{M}_{(\tau)} \times \mathscr{V}$ by

(8.10) $$\Psi(M, V) = \sum_{i,j,k} m_{kj} r_i \lambda_{ij} v_{ki} - \sum_{i,k} r_i \lambda'_{ik} v_{ki} .$$

Then Ψ is a bilinear function of M and V; and the factors $\mathscr{M}_{(\tau)}$ and $\mathscr{V}$ of its domain are both closed, bounded and convex sets in $(n' \times n)$- and $(n' \times m)$-spaces, respectively. Therefore, by a saddle point theorem—see, e.g., Karlin [**9**, (II, theorem 1.51)], there exists a pair $[m^0_{kj}] \equiv M^0 \in \mathscr{M}_{(\tau)}$ and $V^0 \in \mathscr{V}$ such that

(8.11) $$\Psi(M, V^0) \leqq \Psi(M^0, V^0) \leqq \Psi(M^0, V) ,$$

for all $M \in \mathscr{M}_{(\tau)}$ and $V \in \mathscr{V}$.

Suppose (P) holds. Applying our lemma to $V = V^0$, we see that there exists a matrix $M \in \mathscr{M}_{(\tau)}$ such that $\Psi(M, V^0) \geqq 0$. Therefore by (8.11) $\Psi(M^0, V^0) \geqq 0$ and thus

(8.12) $$\Psi(M^0, V) \geqq 0 , \qquad \text{all } V \in \mathscr{V} .$$

Define $V^{(ki)} \in \mathscr{V}$ as a matrix whose (k, i)-th element is 1 and all other elements are 0. Then by (8.10), (8.12)

$$\Psi(M^0, V^{(ki)}) = \sum_j m^0_{kj} r_i \lambda_{ij} - r_i \lambda'_{ik} \geqq 0 ,$$

and since all $r_i > 0$ by (3.1.1), we have

(8.13) $$\sum_i m^0_{kj} \lambda_{ij} \geqq \lambda'_{ik} , \qquad i = 1, \cdots, m; k = 1, \cdots, n' .$$

If in (8.13) at least one of the inequalities is strict, we obtain a contradiction

(8.14) $$\sum_{i,j,k} m^0_{kj} \lambda_{ij} = \sum_{i,j} \lambda_{ij} = m > \sum_{i,k} \lambda'_{ik} = m ,$$

since $\sum_k m^0_{kj} = \sum_j \lambda_{ij} = \sum_k \lambda'_{ik} = 1$. Therefore

(8.15) $$\sum_j m^0_{kj}\lambda_{ij} = \lambda'_{ik}\,, \qquad i = 1, \cdots, m;\ k = 1, \cdots, n', \text{ i.e.,}$$

(8.16) $$\Lambda' = \Lambda M^0_{(\tau)}$$

where $M^0_{(\tau)}$ is the transposed matrix of M^0, hence

(8.17) $$M^0_{(\tau)} \in \mathscr{M}\,.$$

Now put $B = M^0_{(\tau)}$. Then B fulfills the conditions (6.4). This proves that (P) $\Rightarrow$ (B) and completes the proof of our theorem.

We shall now show that the following condition (which will prove useful in Section 9) is equivalent to (B) and thus to (P) and (Λ):

Condition (θ): There exists a matrix $\theta \equiv [\theta_{jk}]$ such that

$$(\theta_0):\ \theta \in \mathscr{M}_{(\tau)}$$
$$(\theta_1):\ \Pi' = \Pi\theta \text{ and}$$
$$(\theta_2):\ q = \theta q'\,.$$

Note that (θ) is a property of the joint distributions P, P', while (B) is a property of the likelihoods Λ, Λ' only.

THEOREM 8.2. $(B) \Leftrightarrow (\theta)$.

PROOF. By definition (see Section 3),

(8.19) $$q_j = \sum_i r_i\lambda_{ij} > 0, \quad q'_k = \sum_i r_i\lambda'_{ik} > 0$$

(8.20) $$q_j\pi_{ij} = r_i\lambda_{ij} \geqq 0, \quad q'_k\pi'_{ik} = r_i\lambda'_{ik} \geqq 0\,.$$

Suppose (B) is true: that is, there exists $B \equiv [\beta_{jk}]$ such that

(8.21) $$\lambda'_{ik} = \sum_j \lambda_{ij}\beta_{jk}; \quad \beta_{jk} \geqq 0, \quad \sum_k \beta_{jk} = 1\,.$$

Then by (8.19)

(8.22) $$q'_k = \sum_{i,j} r_i\lambda_{ij}\beta_{jk} = \sum_j q_j\beta_{jk}\,.$$

Given the matrix B define a matrix $[\theta_{jk}] \equiv \theta$ by the following equations:

(8.23) $$\theta_{jk}q'_k = \beta_{jk}q_j\,, \qquad \text{all } j, k\,.$$

It is easily verified that θ is in $\mathscr{M}_{(\tau)}$ (using (8.22)); that $q = \theta q'$ (using (8.4) and (8.23)); and that $\Pi' = \Pi\theta$ (using (8.20), (8.21), (8.23)). Hence (B) $\Rightarrow$ (θ). To prove the converse, note that condition (θ_1) implies, by (8.20)

$$\pi'_{ik} = \frac{r_i\lambda'_{ik}}{q'_k} = \sum_j \pi_{ij}\theta_{jk} = \sum_j \frac{r_i\lambda_{ij}}{q_j}\theta_{jk}\,,$$

(8.24) $$\lambda'_{ik} = q'_k \sum_j \frac{\theta_{ik}}{q_j}\lambda_{ij}\,.$$

Define $B \equiv [\beta_{jk}]$ by (8.23). Then by (θ_2)

$$\sum_k \beta_{jk} = \frac{1}{q_j}\sum_k \theta_{jk}q'_k = \frac{q_j}{q_j} = 1; \quad \beta_{jk} \geqq 0\,,$$

hence $B \in \mathscr{M}$; moreover by (θ_1) and (8.23), $\Lambda' = \Lambda B$. Thus B satisfies both conditions in (6.4), and we have proved that $(\theta) \Rightarrow$ (B). This completes the proof.

Combining Theorems 8.1 and 8.2, we have

THEOREM 8.3. $(P) \Leftrightarrow (\Lambda) \Leftrightarrow (B) \Leftrightarrow (\theta)$.

Using the matrix $\Gamma' \equiv [p(y \mid y')]$ defined in (3.3′) we introduce

Condition (Γ'): $\Pi' = \Pi\Gamma'$.

Then we have

THEOREM 8.4. $(G) \Rightarrow (\Gamma') \Rightarrow (\theta)$.

PROOF. Consider the identities

$$p(y, z \mid y') = p(y, y', z)/p(y') = p(z \mid y, y')\cdot p(y, y')/p(y')$$
$$p(y, z \mid y') = p(z \mid y, y')\cdot p(y \mid y')\ ;$$

summing over y

$$p(z \mid y') = \sum_y p(z \mid y, y')\cdot p(y \mid y')\ .$$

This becomes, if (G_2) holds,

$$p(z \mid y') = \sum_y p(z \mid y)\cdot p(y \mid y')$$

$$\Pi' = \Pi\Gamma'\ .$$

This proves $(G) \Rightarrow (\Gamma')$. It is clear that if (Γ') holds, then (θ) is satisfied with $\theta = \Gamma'$. This completes the proof.

If a matrix B, and therefore also matrix θ, exists, it is possible to interpret the information systems Y and Y' *as if* Y were garbled into Y', although it is not known whether or not the trivariate distribution on $Z \times Y \times Y'$ satisfies condition (G).

Example. As in Section 2, let z be the visibility at the airport. Let y be the true atmospheric pressure, and y' be the reading on a barometer. Then presumably condition (G) is satisfied, and by the proof of Theorems 6.5 and 8.4, conditions (B) and (θ) are satisfied by matrices $B = \Gamma$ and $\theta = \Gamma'$ (and possibly also other matrices). If, on the other hand, we interpret y', not as a reading on a barometer, but as, say, the true level of humidity, then we have no reason to suppose that Y is garbled into Y' (or conversely). If the bivariate distributions on $Z \times Y$ and $Z \times Y'$, respectively, happen to satisfy condition (B), hence (θ), it may be useful to interpret the matrix θ *as if* it were identical with Γ' (and B with Γ). That is, we can conceive of the messages y_j as follows: given y'_k, the message y_j will be produced, using a random device, with probability $p(y_j \mid y'_k) = \theta_{jk}$, $j = 1, \cdots, n$.

9. COMPARATIVE INFORMATIVENESS REVEALED BY POSTERIOR PROBABILITIES ONLY: THE CASE OF INDEPENDENT Π MATRIX

Let π_j and π'_k be the j-th column of Π and the k-th column of Π' respectively. Now, π_j and π'_k can be interpreted as m-dimensional vectors or points

in m-dimensional space. Accordingly Π and Π' can be interpreted—whenever it is convenient—as two sets of vectors: $\Pi \equiv \{\pi_1, \cdots, \pi_n\}$, $\Pi' \equiv \{\pi'_1, \cdots, \pi'_{n'}\}$. Let S^{m-1} be the simplex defined by the set of all points $v = (v_1, \cdots, v_m)$ such that $v_i \geqq 0$, $\sum v_i = 1$. Clearly both Π and Π' are subsets of S^{m-1}.

Define the following condition: the set Π' is contained in the convex hull $K(\Pi)$ of the set Π:

Condition (K): $\Pi' \subset K(\Pi)$.

This is represented by Figures 2a, 3a, 3b, 4, but not by Figures 2b, 3d. Clearly (K) is equivalent to the following condition: there exists a matrix M such that

$$M \in \mathscr{M}_{(\tau)};\ \Pi' = \Pi M: \tag{9.1}$$

(9.1) is identical with conditions (θ_0) and (θ_1) of Section 8. Hence (θ) which requires, in addition, (θ_2), implies but is not implied by (K). We have thus

THEOREM 9.1. $(\theta) \mapsto (K)$.

In this section we shall consider the case when the columns of the matrix Π (but not necessarily of Π') are linearly independent, that is:

Condition (I_π): (rank of $\Pi) = n$.

This condition will be later shown to apply in two important cases: the binomial case ($n = 2$) of Section 10 and the noiseless case of Section 11.

Similarly we introduce

Condition (I_Λ): (rank of $\Lambda) = n$.

Noting that the r_i and q_j are all > 0 it is easy to prove

THEOREM 9.2. $(I_\pi) \Leftrightarrow (I_\Lambda)$.

Now we shall prove

THEOREM 9.3. *If conditions* (I_π) *and* (θ) *(hence also condition* (B)*) are true, then* (B) *and* (θ) *are satisfied by a unique pair of matrices* B, θ *related by the equations*

$$\theta_{jk} q'_k = \beta_{jk} q_j\,, \qquad \textit{all } j, k\,. \tag{8.23}$$

PROOF. By the hypothesis, we have

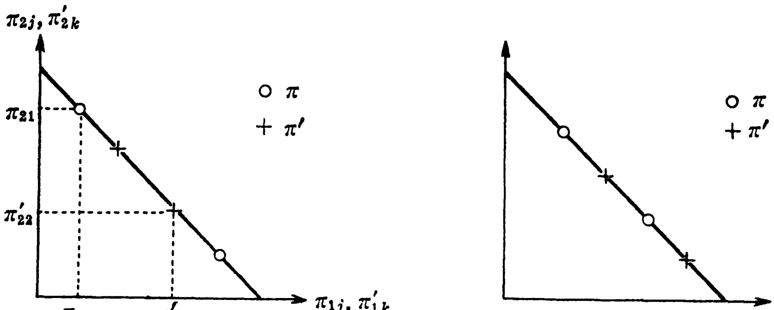
$$(\theta_1): \Pi' = \Pi\theta, \quad (\mathrm{B}_1): \Lambda' = \Lambda B\,,$$

and Π consists of linearly independent columns. Then by Theorem 9.2, Λ also consists of linearly independent columns. Therefore, in the above relations (θ_1) and (B_1), θ and B are unique. Now in proving Theorem 8.2 we have shown that if a matrix B satisfying condition (B) exists then a matrix θ defined by equations (8.23) will satisfy condition (θ); and conversely. Hence the unique pair B, θ satisfying (B), (θ) respectively must also satisfy equations (8.23). Stronger than Theorem 9.3 is

THEOREM 9.4. $[(I_\Lambda)$ *and* $(B)] \Leftrightarrow [(B)$ *with unique* $B] \Leftrightarrow [(\theta)$ *with unique* $\theta]$.

PROOF. By Theorems 9.2 and 9.3, it is sufficient to prove that, when (B)

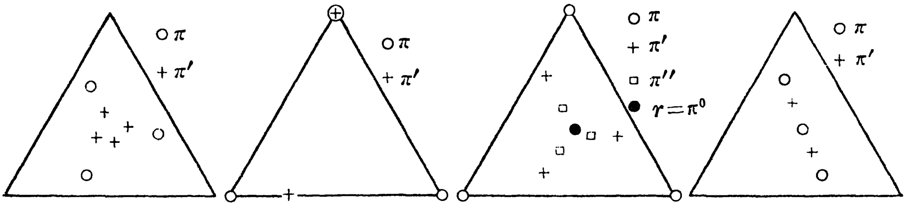
FIGURES 2, 3, 4
SIMPLEX REPRESENTATION OF POSTERIOR PROBABILITIES



2a: $\pi' \subset K(\pi)$, $\pi > \pi'$

2b: $\pi' \not\subset K(\pi)$, $\pi \not> \pi'$

FIGURE 2
$m = 2$
Y, Y' BINARY: $m = 2 = n = n'$



3a: $m=3=n$
$\pi' \subset K(\pi)$, $\pi > \pi'$

3b: $m=3=n$, $n'=2$
Y, Y' noiseless

3c: $m=3$
$\pi > \pi' > \pi'' > \pi^0$

3d

FIGURE 3
$m = 3$

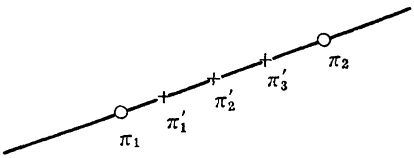


FIGURE 4
$m \geqq 2$
Y BINOMIAL: $n = 2$

holds and B is unique, then ($\mathrm{I_A}$) holds. Suppose (B) holds, with unique $B \equiv [\beta_{jk}]$ but ($\mathrm{I_A}$) does not hold. Let $g < n$ and let $\lambda_1, \cdots, \lambda_g$ be linearly independent, while $\lambda_{g+1}, \cdots, \lambda_n$ are linearly dependent on $\{\lambda_1, \cdots, \lambda_g\}$. That is

$$\lambda_h = \sum_{l=1}^{g} a_{hl}\lambda_l \, , \qquad h = g + 1, \cdots, n \, . \tag{9.2}$$

By condition (B) and (9.2)

$$\lambda'_k = \sum_{j=1}^{n} \beta_{jk}\lambda_j = \sum_{l=1}^{g}\left(\beta_{lk} + \sum_{h=g+1}^{n} a_{hl}\beta_{hk}\right)\lambda_l , \qquad k = 1, \cdots, n' . \tag{9.3}$$

Let us define a matrix $\Delta \equiv [\delta_{jk}]$ as follows:

$$\delta_{j1} = \beta_{j1} + \varepsilon_j \tag{9.4}$$

$$\delta_{js} = \beta_{js} - \frac{1}{n'-1}\varepsilon_j , \qquad s = 2, \cdots, n'; j = 1, \cdots, n , \tag{9.5}$$

where ε_j, $j = 1, \cdots, n$ are numbers such that

$$\varepsilon_l + \sum_{h=g+1}^{n} a_{hl}\varepsilon_h = 0 , \qquad l = 1, \cdots, g . \tag{9.6}$$

We shall prove our theorem for the case in which all $\beta_{jk} > 0$. Clearly it is possible to choose all ε_j, $j = 1, \cdots, n$ which satisfy (9.6) and have arbitrarily small absolute values. Then, for $\beta_{jk} > 0$ and $|\varepsilon_j|$ sufficiently small, we have by (9.4), (9.5)

$$\delta_{jk} \geqq 0, \qquad \sum_k \delta_{jk} = \sum_k \beta_{jk} = 1 . \tag{9.7}$$

By (9.4), (9.5), (9.6),

$$\delta_{lk} + \sum_h a_{hl}\delta_{hk} = \beta_{lk} + \sum_h a_{hl}\beta_{hk} . \tag{9.8}$$

Therefore by (9.2), (9.3), (9.8),

$$\lambda'_k = \sum_{l=1}^{g}\left(\delta_{lk} + \sum_h a_{hl}\delta_{hk}\right)\lambda_l = \sum_{j=1}^{n} \delta_{jk}\lambda_j . \tag{9.9}$$

Then it follows from (9.7) and (9.8) that $\Delta \in \mathscr{M}$, $\Delta \neq B$, and $\Lambda' = \Lambda\Delta$. The part of the theorem involving θ is proved similarly.

We shall now show that, under condition (I_π), if Y' is obtained by garbling Y—see Section 6—then the matrices B and θ are identical, respectively, with the matrices $\Gamma \equiv [\gamma_{jk}]$ and $\Gamma' \equiv [\gamma'_{jk}]$ defined in Section 3:

$$\gamma_{jk} \equiv p(y'_k \mid y_j); \qquad \gamma'_{jk} \equiv p(y_j \mid y'_k) .$$

THEOREM 9.5. $[(I_\pi)$ *and* $(G)] \Rightarrow (B = \Gamma, \theta = \Gamma')$.

PROOF. By Theorem 6.5 and 8.2, condition (G) implies that

$$\Lambda' = \Lambda\Gamma, \quad \Lambda' = \Lambda B, \quad \Pi' = \Pi\theta .$$

Hence $\Lambda(\Gamma - B) = 0$. Let (I_π) hold. Then by Theorem 9.2, Λ consists of independent columns, and therefore $\Gamma = B$, i.e.:

$$\gamma_{jk} \equiv p(y'_k \mid y_j) = \beta_{jk} , \qquad \text{all } j, k .$$

By Theorem 9.3, θ is uniquely defined by

$$\begin{aligned}\theta_{jk} &= \beta_{jk}q_j/q'_k = p(y'_k \mid y_j)p(y_j)/p(y'_k)\\ &= p(y_j \mid y'_k) ;\end{aligned}$$

hence

$$[\theta_{jk}] = [\gamma'_{jk}], \qquad \theta = \Gamma' .$$

We shall now prove that (Γ) does not imply (G), nor does (B) imply (Γ). Condition (P) means

$$p(y' \mid z) = \sum_y p(y \mid z) p(y' \mid y) . \tag{9.10}$$

On the other hand we have identically

$$p(y' \mid z) = \sum_y p(y, y' \mid z) = \sum_y p(y \mid z) p(y' \mid z, y) . \tag{9.11}$$

If Π is linearly dependent, then by Theorem 9.4, equations (9.10), (9.11) are consistent with the inequality

$$p(y' \mid y) \neq p(y' \mid z, y) ,$$

as contrasted with (G_1). Hence (G) does not necessarily follow from (Γ). Similarly we can prove that (B) does not imply (Γ).

We shall now show that the independence condition (I_π) is both sufficient and necessary for the equivalence of the conditions (P), (Π) and (K). We shall start with

THEOREM 9.6. $(I_\pi) \Rightarrow ((K) \Leftrightarrow (P))$.

PROOF. By Theorems 8.3 and 9.1, we always have (P) ⇒ (K). Therefore we need only prove that (K) ⇒ (P) if the linear independence condition (I_π) holds; i.e., that

$$(\mathrm{I}_\pi) \Rightarrow ((\mathrm{K}) \Rightarrow (\mathrm{P})) . \tag{9.12}$$

Let Y and Y' be two information systems with known $P^Y \equiv (\Pi_*, q_*)$ and $P^{Y'} \equiv (\Pi'_*, q'_*)$ respectively, where by (3.2)

$$\Pi_* q_* = \Pi'_* q'_* . \tag{9.13}$$

Suppose (K) holds; that is, there exists a matrix M satisfying (9.1) with $\Pi = \Pi_*$, $\Pi' = \Pi'_*$. Then by (9.13)

$$\Pi_*(q_* - M q'_*) = 0 . \tag{9.14}$$

If (I_π) holds for Π_*, then, the linear independence of its columns implies

$$q_* = M q'_* , \tag{9.15}$$

so that, by (9.1), all three components of condition (θ) hold. Hence by Theorem 8.3, (P) holds. This proves that

$$(\mathrm{I}_\pi) \Rightarrow ((\mathrm{K}) \Rightarrow (\mathrm{P})) . \tag{9.16}$$

We shall now prove

THEOREM 9.7. $(I_\pi) \Rightarrow ((P) \Leftrightarrow (\Pi))$.

PROOF. Since by Theorem 5.1, (Π) always implies (P), we need only to prove that (P) implies (Π) if (I_π) holds; i.e., that

$$(I_\pi) \Rightarrow ((P) \Rightarrow (\Pi)) . \tag{9.17}$$

First note that since (P) is equivalent to (θ) by Theorem 8.3, condition (Π) is by definition equivalent to the following: for any q, q' such that

$$\Pi_* q = \Pi'_* q' \tag{9.18}$$

there exists a matrix θ such that

$$\begin{aligned} \theta \quad &\in \mathscr{M}_{(\tau)} , \\ \Pi'_* &= \Pi_* \theta , \\ q \quad &= \theta q' . \end{aligned} \tag{9.19}$$

The last three conditions correspond to the three components of condition (θ), applied here to all q, q' consistent with (9.18), and not only (as in (9.13)) to a particular q_*, q'_*.

Suppose (P), or equivalently (θ), holds for the known (starred) distributions; that is, there exists a matrix θ_* (say) such that

$$\theta_* \in \mathscr{M}_{(\tau)} , \tag{9.20.0}$$

$$\Pi'_* = \Pi_* \theta_* , \tag{9.20.1}$$

$$q_* = \theta_* q'_* . \tag{9.20.2}$$

Thus, under condition (P), there exists at least one triple (q, q', θ) which satisfied (9.19), viz., the triple (q_*, q'_*, θ_*). Now let q, q' be undetermined except for the constraint (9.18). Then by (9.20.1),

$$\Pi_*(q - \theta_* q') = 0 . \tag{9.21}$$

If the columns of Π_* are linearly independent then (9.21) implies $q = \theta_* q'$; i.e., (P) then implies (Π) with $\theta = \theta_*$. Therefore

$$(I_\pi) \Rightarrow ((P) \Rightarrow (\Pi)) . \tag{9.22}$$

This completes the proof.

Combining Theorems 9.6 and 9.7, we have

THEOREM 9.8. $(I_\pi) \Rightarrow ((K) \Leftrightarrow (P) \Rightarrow (\Pi))$.

Example. Let $m = n = n' = 3$. Figure 3c shows a sequence of triangles (with vertices representing the matrices Π, Π', Π'', Π^0), nestling in each other and containing the point r; the largest triangle coincides with the simplex itself, and the sequence shrinks to the point r as its lower bound. The vertices of any of these triangles represent the set $\Pi = \{\pi_1, \pi_2, \pi_3\}$ of some information system Y, and the triangle itself is $K(\Pi)$. If the triangle corresponding to Y' is nestled in the triangle corresponding to Y, then $\Pi' \subset K(\Pi)$; and since the vertices are not colinear, condition (I_π) is satisfied. Hence by Theorem 9.2, $\Pi > \Pi'$: the system Y is more informative than Y', as revealed by the posterior probabilities alone. Note that the vertices of the largest triangle in the sequence, i.e. of the simplex itself, give perfect information; while the lower bound represented by point r is just as informative as "null information," i.e. the knowledge of the prior probabilities alone, without any messages. The sequence is ordered according to informativeness, and so is any other sequence of nestling triangles. The set of all triangles is a lattice, partially ordered by the relation "more informative than," with "complete information" as the maximum element, and "null information" as the infimum.

10. BINOMIAL AND BINARY INFORMATION SYSTEMS

The numbers m, n, n' are always $\geqq 2$. Following Blackwell and Girshick [4] the case $m = 2$ will be called *dichotomy*. If information system Y consists of 2 messages, $n = 2$, it is called *binomial*; and when $m = n = 2$, the information—theoretical literature—e.g., Feinstein [7]—calls the channel (the information system) Y, *binary*. We shall denote these conditions thus:

$$\text{(d)}:\quad m = 2; \qquad \text{(b)}:\quad n = 2; \qquad \text{(b}'\text{)}:\quad n' = 2\,.$$

Thus a binary channel is defined by (b) and (d).

Since always $m \geqq 2$, condition (b) implies that the columns of the matrix Π are linearly independent. Thus (b) implies (I_π), but of course, not conversely. Hence by Theorem 9.8

THEOREM 10.1. $(b) \mapsto ((K) \Leftrightarrow (\Pi) \mapsto (P))$.

The case is illustrated on Figure 4, where the points π_1, π_2 representing Y, and the points π_1', π_2', π_3' representing Y' are all arranged on a straight line in a space of m dimensions (arbitrary $m \geqq 2$), and all π_k' lie *between* π_1 and π_2.

The more special case $n = n' = m = 2$—i.e. (b), (b′) and (d)—occurs in the testing of simple hypotheses and in comparing "binary channels." After labelling the two events z_1 and z_2 arbitrarily, let us label the two messages in Y so that

$$\pi_{11} \geqq \pi_{12}\,. \tag{10.1}$$

Then condition (K) takes the form

$$\begin{aligned} \pi_{11} &\geqq \pi_{11}' \geqq \pi_{12}\,, \\ \pi_{11} &\geqq \pi_{12}' \geqq \pi_{12}\,, \end{aligned} \tag{10.2}$$

and by Theorem 10.1, we have

THEOREM 10.2. *Let Y and Y' be two binary channels* (*i.e., let $m = n = n' = 2$*); *then Y is more informative than Y' if and only if the posterior probabilities satisfy the relation* (10.2); *provided the two messages in Y are labelled so as to make $\pi_{11} \geqq \pi_{12}$.*

This criterion (usable only when the posterior probabilities are known) is simpler than the one provided by Blackwell-Girshick [4, (section 12.5)] and based on the likelihood matrix Λ.

11. COMPARISON OF NOISELESS INFORMATION SYSTEMS

Let Y be a noiseless information system, i.e. $Z \mathrm{s} Y$ (see Section 2). Then $\lambda_{ij} \equiv p(y_j \mid z_i) = 1$ or 0, all i, j. In general, $\sum_j \lambda_{ij} = 1$ and $\sum_i \gamma_i \lambda_{ij} = q_j > 0$. Therefore, if Y is noiseless, then each row of its Λ matrix contains one and only one non-zero ($= 1$) element and each column contains at least one non-zero ($= 1$) element. Now $\pi_{ij} = r_i \lambda_{ij} / q_j$. Hence, whenever $\lambda_{ij} = 1$ or 0, then $\pi_{ij} = r_i / q_j$ (> 0) or 0, respectively. Therefore;

(11.1) *If Y is noiseless, then each row of its Π matrix contains one and only one non-zero* (> 0) *element and each column contains at least one non-zero* (> 0) *element.*

Using (11.1) we shall prove the following property of condition (N) defined in Section 3:

THEOREM 11.1. $(N) \mapsto ((I_\pi)$ *and* $(I_{\pi'}))$.

PROOF. Suppose Y is noiseless and let

$$\sum_{j=1}^{n} g_j \pi_{ij} = 0 \, , \qquad \text{all } i = 1, \cdots, m \, , \tag{11.2}$$

for some real numbers $g_1, \cdots, g_n$. We shall then show that $g_j = 0$, for all j, i.e., the columns of Π are linearly independent. We note that by (11.1), for any fixed j, say $j = j_1$, there is at least one i, say $i = i_1$, such that $\pi_{i_1 j_1} > 0$. Then by (11.1), $\pi_{i_1 j} = 0$, for all $j \neq j_1$. Therefore, for $i = i_1$, (11.2) becomes $g_{j_1}\pi_{i_1 j_1} = 0$. Since $\pi_{i_1 j_1} > 0$, we have

$$g_{j_1} = 0 \, , \qquad j_1 = 1, \cdots, n \, ; \tag{11.3}$$

that is, (I_π) holds. Similarly, for Y' noiseless, we have $(I_{\pi'})$. The converse obviously does not hold.

Using Theorems 9.7 and 11.1, we have

THEOREM 11.2. $(N) \mapsto ((P) \Leftrightarrow (\Pi) \Leftrightarrow (K))$.

In Section 2, we defined condition (C); and by Theorem 7.3, $(C) \mapsto (P)$. Hence by Theorem 11.2, we have

THEOREM 11.3. $(N) \mapsto ((C) \Rightarrow (\Pi) \Leftrightarrow (P) \Leftrightarrow (K))$.

Thus, if both Y and Y' are noiseless, and Y is finer than Y', then Y is more informative than Y', as revealed by posterior probabilities of events

$$\Pi' = \Pi M, \; M \equiv [\mu_{jk}] \in \mathscr{M}_{(r)} \, , \tag{11.4}$$

which is, of course, our condition (K).

Remark 1. The matrix $M \equiv [\mu_{jk}]$ in (11.4), with $Z \mathrm{s} Y \mathrm{s} Y'$, $\Pi' = \Pi M$, has the following interpretation:

$$M = \Gamma' \equiv [\gamma'_{jk}] \equiv [p(y_j \mid y'_k)] \, ,$$

since by Theorems 11.1, 11.3, 7.1, 9.5, we have $\theta = \Gamma' = M$.

Example. Let $Z \mathrm{s} Y \mathrm{s} Y'$, so that by Theorem 11.3 and Remark 1, $\Pi' = \Pi M$, $\mu_{jk} = p(y_j \mid y'_k)$; let the posterior probability matrices be as follows (with $0 < \pi'_{11} = 1 - \pi'_{21} < 1$);

$$\begin{array}{ccc} \Pi & \qquad & \Pi' \\ \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} & & \begin{matrix} \pi'_{11} & 0 \\ \pi'_{21} & 0 \\ 0 & 1 \, . \end{matrix} \end{array}$$

Then $y'_1 = y_1 \cup y_2$, $y'_2 = y_3$. Since Π is the identity matrix I, we have $\Pi' = \Pi M = IM = M = \Gamma'$. The case is illustrated in Figure 3b, where the point π'_1 divides the base of the triangle in proportion $\pi'_{21} : \pi'_{11} = p(y_2 | y'_1) : p(y_1 | y'_1)$.

Now we shall prove the following

THEOREM 11.4. $(N) \models ((C) \Leftrightarrow (K) \Leftrightarrow (\Pi) \Leftrightarrow (P))$, *when* X *is finite.*

PROOF. By Theorems 11.2 and 11.3, it will suffice to prove that

$$(\mathrm{N}) \models ((\mathrm{K}) \Rightarrow (\mathrm{C})), \text{ when } X \text{ is finite.}$$

Let us assume (N) and (K): that is, $Z\mathrm{s}Y$, $Z\mathrm{s}Y'$ and (11.4) holds, so that

$$\pi'_{ik} = \sum_j \pi_{ij}\mu_{jk} , \qquad \text{all } j, k . \tag{11.5}$$

For any fixed k, say $k = k_1$, let $z_{i_1} \in Z$ be contained in y'_k (remember that Z is a sub-partition of the noiseless information system Y'.) Then $\lambda'_{i_1k_1} = 1$, and

$$\pi'_{i_1k_1} > 0 . \tag{11.6}$$

Since Y is noiseless, by (11.1) the i_1-th row of Π contains one and only one non-zero element, say $\pi_{i_1j_1} > 0$. Then

$$\pi_{i_1j_1} > 0, \qquad \pi_{i_1j} = 0 , \qquad \text{for all } j \neq j_1 . \tag{11.7}$$

In (11.5), let $i = i_1$ and $k = k_1$; then by (11.7),

$$\pi'_{i_1k_1} = \pi_{i_1j_1}\mu_{j_1k_1} ; \tag{11.8}$$

hence

$$\mu_{j_1k_1} > 0 , \tag{11.9}$$

by (11.6), (11.7), (11.8). Suppose X is finite. Then $\pi_{i_1j_1} > 0$, (i.e., $\lambda_{i_1j_1} = 1$) implies $z_{i_1} \subset y_{j_1}$; therefore, (11.7) admits of the following interpretation:

(11.10) *For any fixed* $k = k_1$, *and any* $z_{i_1} \subset y'_{k_1}$, *there exists at least one* y_{j_1} *such that* $z_{i_1} \subset y_{j_1}$.

Now, for this y_{j_1}, let $z_{i'} \in Z$ be contained in y_{j_1}. Then $\lambda_{i'j_1} = 1$ and

$$\pi_{i'j_1} > 0, \qquad \pi_{i'j} = 0 , \qquad \text{for all } j \neq j_1 , \tag{11.11}$$

by (11.1). Then, for $i = i'$ and $k = k_1$ (11.5) becomes

$$\pi_{i'k_1} = \pi_{i'j_1}\mu_{j_1k_1} , \tag{11.12}$$

so that by (11.9), (11.11) and (11.12), $\pi_{i'k_1} > 0$, $\mu_{i'k_1} = 1$. When X is finite this implies $z_{i'} \subset y'_{k_1}$. This proves that

(11.13) *For any* y_{j_1} *such that* $y_{j_1} \cap y'_{k_1} \neq \emptyset \quad y_{j_1} \subset y'_{k_1}$.

From (11.10) and (11.13), we conclude that, if (N) and (K) hold and X is finite, then any y'_k is a union of disjoint y's, i.e., $Y\mathrm{s}Y'$, viz., (C) holds. This completes the proof.

Remark 2. Let Y be noiseless: If the columns of Π' are linear combinations of Π, that is, if there exists a $n \times n'$ matrix $A = [\alpha_{jk}]$ such that $\Pi' = \Pi A$, i.e.,

$$\pi'_{ik} = \sum_j \pi_{ij}\alpha_{jk} , \qquad \text{all } i, k , \tag{11.14}$$

then $A \in \mathscr{M}^{(c)}_{n\times n'}$; hence (K) holds.

To prove, sum both sides of (11.14) with respect to i; then $\sum_j \alpha_{jk} = 1$, all k (since $\sum_i \pi'_{ik} = \sum_i \pi_{ij} = 1$). It remains to prove that $\alpha_{jk} \geqq 0$ for all j, k. Fix $j = j_1$, $k = k_1$, and let $i = i_1$ be such that $\pi_{i_1 j_1} > 0$. Since Y is noiseless, we have by (11.1), (11.14)

$$\pi'_{i_1 k_1} = \pi_{i_1 j_1} \alpha_{j_1 k_1} .$$

Since $\pi_{i_1 j_1} > 0$, if $\pi'_{i_1 k_1} > 0$ we have $\alpha_{j_1 k_1} > 0$; and if $\pi'_{i_1 k_1} = 0$, then $\alpha_{j_1 k_1} = 0$. Hence $\alpha_{jk} \geqq 0$, all j, k, completing the proof.

12. CONVEX OPERATORS ON POSTERIOR PROBABILITIES: THE EQUIVOCATION PARAMETER

For any payoff function ω, we shall define a function υ_ω on the simplex S^{m-1} (i.e., on the set of all vectors $\xi \equiv (\xi_1, \cdots, \xi_m)$ such that $\xi_i \geqq 0$, $\xi_i = 1$) as follows:

$$\upsilon(\xi) \equiv \max_{d \in D^\omega} \sum_{i=1}^{m} \omega(z_i, d)\xi_i . \tag{12.1}$$

Then by (4.3), we have

$$U(P; \omega) \equiv \sum_{j=1}^{n} q_j \upsilon_\omega(\pi_j) , \tag{12.2}$$

where π_j is the j-th column of Π. Consider a pair of probability distributions (Π, q), (Π', q') with the following property:[13]

Condition (φ): For any convex function φ on S^{m-1}

$$\sum_{k=1}^{n} q_j \varphi(\pi_j) \geqq \sum_{k=1}^{n'} q_k \varphi(\pi'_k) . \tag{12.3}$$

We shall prove[14]

THEOREM 12.1. $(P) \Leftrightarrow (\varphi)$.

PROOF. First we shall prove $(\varphi) \Rightarrow$ (P). We have by (12.1), for any two vectors $\xi_1 \equiv (\xi_{11}, \cdots, \xi_{m1})$, $\xi_2 \equiv (\xi_{12}, \cdots, \xi_{m2})$, both in S^{m-1}, and letting $0 \leqq \alpha \leqq 1$,

$$\begin{aligned} \upsilon_\omega(\alpha\xi_1 + (1-\alpha)\xi_2) &\leqq \alpha \max_{d \in D^\omega} \sum_i \omega(z_i, d)\xi_{i1} + (1-\alpha) \max_{d \in D^\omega} \sum_i \omega(z_i, d)\xi_{i2} \\ &= \alpha\upsilon_\omega(\xi_1) + (1-\alpha)\upsilon_\omega(\varphi_2) . \end{aligned} \tag{12.4}$$

That is, υ_ω is a convex function on S^{m-1}. Therefore, using expression (12.2) for $U(P; \omega)$ and a similar one for $U(P'; \omega)$, condition (φ) is seen to imply that $U(P; \omega) \geqq U(P'; \omega)$, all ω, i.e., condition (P).

[13] In DeGroot's [6] definition of an information amount $I[Y; r; \phi]$ it is reasonable to require that $I[Y, r; \phi] \geq 0$ for all Y and r. He proves that an (uncertainty) function ϕ is concave on S^{m-1} is a necessary and sufficient condition for that.

[14] Compare also Blackwell-Girshick [4]. While that book is restricted to finite sets of messages and of events, earlier work of Blackwell [2, 3] extends also to cases when the set of messages are continuous. Accordingly our conditions (B), (W), (φ), all equivalent to (P), are generalized into $(\tilde{\text{B}})$, $(\tilde{\text{W}})$, $(\tilde{\varphi})$, say. Blackwell [3] proved that $(\tilde{\text{B}}) \Rightarrow (\tilde{\text{W}})$ and also $(\tilde{\text{B}}) \Rightarrow (\tilde{\varphi})$; compare also DeGroot [6]. We do not know whether, conversely, corresponding to our theorems for finite information systems, one also has $(\tilde{\text{W}}) \Rightarrow (\tilde{\text{B}})$ and $(\tilde{\varphi}) \Rightarrow (\tilde{\text{B}})$.

Next we shall prove (P) $\Rightarrow$ (φ). By Theorem 8.3, (P) $\Leftrightarrow$ (θ). Therefore, if (P) holds, then there exists

$$\theta = [\theta_{jk}]$$

such that

$$\theta_{jk} \geqq 0, \qquad \sum_j \theta_{jk} = 1 , \tag{12.5.0}$$

$$\pi'_k = \sum_j \pi_j \theta_{jk} , \qquad \text{all } k , \tag{12.5.1}$$

$$q_j = \sum_k \theta_{jk} q'_k , \qquad \text{all } j . \tag{12.5.2}$$

Hence for any convex function φ on S^{m-1}, we have

$$\begin{aligned} \sum_j q_j \varphi(\pi_j) &= \sum_k q'_k \sum_j \theta_{jk} \varphi(\pi_j) \\ &\geqq \sum_k q'_k \varphi\Big(\sum_j \theta_{jk} \pi_j\Big) = \sum_k q'_k \varphi(\pi'_k) ; \end{aligned} \tag{12.6}$$

hence (P) $\Rightarrow$ (φ), completing the proof of the theorem.

Let φ^* be a particular convex function on S^{m-1}, and suppose

Condition (φ^*): $\sum_{j=1}^{n} q_j \varphi^*(\pi_j) \geqq \sum_{k=1}^{n'} q'_k \varphi^*(\pi'_k)$.

Clearly (φ^*) cannot be a sufficient condition for (P) to be satisfied: for (P) induces only a partial ordering on the set of information systems while (φ^*) induces a complete ordering. At the same time, by Theorem 12.1, (φ^*) is a necessary condition for (P). Thus

THEOREM 12.2. $(P) \Leftrightarrow (\varphi) \mapsto (\varphi^*)$.

An important special case is that of the equivocation parameter $H(Z \mid Y)$ of classical information theory, defined in (5.1) or, in another notation, by

$$H(Z \mid Y) \equiv -\sum_j q_j \sum_i \pi_{ij} \log \pi_{ij} . \tag{12.7}$$

The function

$$H(\xi) \equiv \sum_i \xi_i \log \xi_i ,$$

where $\xi \equiv (\xi_1, \cdots, \xi_m) \in S^{m-1}$, is well known to be convex.[15] Therefore, if we introduce

Condition (H): $H(Z \mid Y) \leqq H(Z \mid Y')$.

we can replace, Theorem 12.2, (φ^*) by (H) and obtain

THEOREM 12.3. $(P) \mapsto (H)$:

i.e., lower equivocation is necessary but not sufficient for higher informativeness.

13. SOME RELATIONS WITH BLACKWELL'S RESULTS

Using our terminology and notation we can say that in Blackwell [2], an information system Y is defined by its likelihood matrix Λ without referring

[15] Since $d^2(\xi_i \log \xi_i)/d\xi_i^2 = 1/\xi_i > 0$ for non-negative ξ_i, the matrix of second derivatives of $H(\xi)$, which is diagonal, is positive definite—a suffient condition for convexity.

to the probability distribution on Z. Given an information system Y and a payoff function ω, we shall define a set W_Y^ω in m-dimensional space as follows. Let D^ω be the set of decisions related to ω and let Δ_Y^ω be the set of all decision rules δ from Y to D^ω. Corresponding to each $\delta \in \Delta_Y^\omega$, define a point

$$\omega(\delta) \equiv (\omega_1(\delta), \cdots, \omega_m(\delta)) \tag{13.1}$$

in m-dimensional space by

$$\omega_i(\delta) \equiv \sum_y \omega(z_i, \delta(y))p(y \mid z_i) , \qquad i = 1, \cdots, m . \tag{13.2}$$

Then the set W_Y^ω is defined as follows

$$W_Y^\omega \equiv \{w(\delta); \delta \in \Delta_Y^\omega\} . \tag{13.3}$$

The set $W_{Y'}^\omega$ is defined similarly. Assume that both W_Y^ω and $W_{Y'}^\omega$ are closed sets. In Blackwell's definition,[16] Y is said to be more informative than Y', if the convex hull of W_Y^ω, $K(W_Y^\omega)$, contains $W_{Y'}^\omega$, for all ω. We shall call *Condition* (W): $K(W_Y^\omega) \supset W_{Y'}^\omega$, for all ω.

Blackwell [3] and Blackwell-Girshick [4] have proved

THEOREM 13.1. $(W) \Leftrightarrow (B)$.

From this theorem and our Theorem 8.3, we immedicately obtain

THEOREM 13.2. $(P) \Leftrightarrow (\Lambda) \Leftrightarrow (W) \Leftrightarrow (B) \Leftrightarrow (\theta)$.

Here we shall give a direct proof of the equivalence of three conditions (P), (Λ) and (W) without going through condition (B)

$$(\mathrm{P}) \Leftrightarrow (\Lambda) \Leftrightarrow (\mathrm{W}) . \tag{13.4}$$

First note that $U(P; \omega)$ can be expressed by means of the set W_Y^ω defined in (13.3)

$$U(P; \omega) \equiv U(\Lambda_*, r_*; \omega) = \max_{\delta \in \Delta_Y^\omega} \sum_i r_{*i} \sum_y \omega(z_i, \delta(y))p(y \mid z_i)$$

$$= \max_{\delta \in \Delta_Y^\omega} \sum_i r_{*i} w_i(\delta) , \qquad \text{where } P \equiv (\Lambda_*, r_*) ,$$

$$U(P; \omega) \equiv U(\Lambda_*, r_*; \omega) = \max_{w \in W_Y^\omega} \sum_i r_{*i} w_i , \quad \text{where } w = (w_1, \cdots, w_m) . \tag{13.5}$$

Similarly, we have

$$U(P'; \omega) \equiv U(\Lambda'_*, r_*; \omega) = \max_{w \in W_{Y'}^\omega} \sum_i r_{*i} w_i , \qquad \text{where } P' \equiv (\Lambda'_*, r_*) . \tag{13.6}$$

Suppose (W) holds, i.e., $K(W_Y^\omega) \supset W_{Y'}^\omega$, for all ω. Then for any r,

$$\max_{w \in W_Y^\omega} \sum_i r_i w_i \geqq \max_{w \in W_{Y'}^\omega} \sum_i r_i w_i , \qquad \text{for all } \omega . \tag{13.7}$$

By (13.5) and (13.6), the relation (13.7) is equivalent to

$$U(\Lambda_*, r; \omega) \geqq U(\Lambda'_*, r; \omega) , \qquad \text{for all } r, \omega . \tag{13.8}$$

Hence (W) implies (Λ). And since clearly $(\Lambda) \Rightarrow (\mathrm{P})$, we have proved that

[16] A definition of more informativeness by condition (W) was originally proposed by Bohnenblust, Shapley and Scherman [5].

(W) $\Rightarrow$ (Λ) $\Rightarrow$ (P).

Next, we shall prove that (P) implies (W); that is, if

$$u(\Lambda_* r_*; \omega) \geqq u(\Lambda'_*, r_*; \omega)\,, \qquad \text{for all } \omega\,, \tag{13.9}$$

then (W) holds. (The proof is similar to that of Theorem 8.1.) Let $(g_1, \cdots, g_m)$ be an arbitrary set of m real numbers; define

$$h_i \equiv g_1/r_{*i} \qquad i = 1, \cdots, m\,. \tag{13.10}$$

For any payoff function ω, we define a payoff function $\bar{\omega}$ by $D^{\bar{\omega}} = D^{\omega}$ and

$$\bar{\omega}(z_i, d) = h_i\omega(z_i, d), \qquad d \in D^{\omega},\ i = 1, \cdots, m\,. \tag{13.11}$$

Then it is clear that

$$\bar{w}_i(\delta) \equiv \sum_y \bar{\omega}(z_i, \delta(y))p(y \mid z_i) = h_i w_i(\delta)\,, \qquad i = 1, \cdots, m\,, \tag{13.12}$$

where $\delta \in \Delta_Y^{\omega} \equiv \Delta_Y^{\bar{\omega}}$, and $w_i(\delta)$ is defined by (13.2).

Let us define a linear transformation T from m-space to m-space by $Tw \equiv T(w_1, \cdots, w_m) = (h_1 w_1, \cdots, h_m w_m)$. Then from (13.12) we have $\bar{w}(\delta) \equiv (\bar{w}_1(\delta), \cdots, \bar{w}_m(\delta)) = Tw(\delta)$, for all $\delta \in \Delta_Y^{\omega} \equiv \Delta_Y^{\bar{\omega}}$. Therefore W_Y^{ω} defined by (13.3) and $W_Y^{\bar{\omega}}$ defined similarly with respect to $\bar{\omega}$ are related by

$$W_Y^{\bar{\omega}} = TW_Y^{\omega}\,. \tag{13.3}$$

From (13.5) and (13.13), we have

$$\begin{aligned} U(\Lambda_*, r_*; \bar{\omega}) &= \max_{w \in W_Y^{\bar{\omega}}} \sum r_{*i} w_i \\ &= \max_{w \in W_Y^{\omega}} \sum r_{*i}(Tw)_i \end{aligned} \tag{13.14}$$

where $(Tw)_i$ is the i-th coordinate of the point Tw, that is, from the definition of T and (13.10),

$$(Tw)_i = h_i w_i = \frac{g_i}{r_{*i}} w_i\,, \qquad i = 1, \cdots, m\,. \tag{13.15}$$

Therefore, from (13.14) and (13.15), we have

$$U(\Lambda_*, r_*; \bar{\omega}) = \max_{w \in W_Y^{\omega}} \sum_i g_i w_i\,. \tag{13.16}$$

Similarly

$$U(\Lambda'_*, r_*; \bar{\omega}) = \max_{w \in W_{Y'}^{\omega}} \sum_i g_i w_i\,. \tag{13.16$'$}$$

From (13.9), (13.16) and (13.16$'$) we have

$$\max_{w \in W_Y^{\omega}} \sum_i g_i w_i \geqq \max_{w \in W_Y^{\omega}} \sum_i g_i w_i\,, \tag{13.17}$$

where $(g_1, \cdots, g_m)$ and ω is any set of m real numbers and ω is any payoff function. Therefore (13.17) implies

$$K(W_Y^{\omega}) \supset W_Y^{\omega}\,, \qquad \text{for all } \omega\,,$$

that is, (W). Accordingly (P) $\Rightarrow$ (W) is proved. This completes a proof of

the equivalence relation (13.4) which includes Theorem 5.2 as a part of it.

14. DICHOTOMIES: $m = 2$

In the dichotomy case, Z consists of two events z_1 and z_2: $m = 2$. As in (9.19), let

$$\Pi_* q = \Pi'_* q' (= r) , \tag{14.1}$$

where Π_*, Π'_*, but not q, q', r are fixed. Henceforth we shall omit the asterisk without ambiguity.

With respect to an information system Y, define two functions $G_Y(t)$ and $F_Y(t)$ on $[0, 1]$ as follows:[17]

$$G_Y(t) \equiv \sum_{\pi_{1j} \leq t} q_j , \tag{14.2}$$

$$F_Y(t) \equiv \int_0^t G_Y(u) du . \tag{14.3}$$

$G_{Y'}(t)$ and $F_{Y'}(t)$ are defined similarly with respect to Y'. Clearly $G_Y(t)$ is a monotone non-decreasing step function with a jump of q_j at $t = \pi_{1j}$, $j = 1, \cdots, n$.

From the consistency condition (14.1),

$$F_Y(1) = \int_0^1 G_Y(t) dt = 1 - \sum_j q_j \pi_{1j} = 1 - r_1 = r_2 ,$$

$$F_{Y'}(t) = \int_0^1 G_{Y'}(t) dt = 1 - \sum_k q'_k \pi'_{1k} = 1 - r_1 = r_2 .$$

Therefore

$$F_Y(1) = F_{Y'}(1) = r_2 . \tag{14.4}$$

It will become clear that the relation (14.4) is a key point in our proof of Theorem 14.3 below. Now we shall introduce

Condition (F): $F_Y(t) \geqq F_{Y'}(t)$, all t, $0 \leqq t \leqq 1$.

Then, following a reasoning similar to that of Blackwell-Girshick [4, (theorem 12.4.1)], it is easy to prove[18]

[17] Compare $G_Y(t)$ and $F_Y(t)$ with $F_P(t)$ and $C_P(t)$ in Blackwell-Girshick [4, (theorem 12.4.1)]. If $r_1 = r_2 = 1/2$, then our $G_Y(t), F_Y(t)$ become equal to Blackwell-Girshick's $F_P(t), C_P(t)$ respectively.

[18] We shall show that Theorem 14.1 can be proved by almost the same device as given by Blackwell-Girshick [4] in their proof of their Theorem 12.4.1. When $m = 2$, condition (φ) is equivalent to the following condition: For any convex function φ on $[0, 1]$,

$$\sum_{j=1}^{n} q_j \varphi(\pi_{1i}) \geqq \sum_{k=1}^{n'} q_{k'} \varphi(\pi_{1k}) .$$

Defining a function $f_t(u)$ on $[0, 1]$ for each $t, 0 \leqq t \leqq 1$, by

$$f_t(u) = \begin{cases} t - u, & \text{for } 0 \leqq u \leqq t \\ 0, & \text{for } 0 \leqq u \leqq 1 , \end{cases}$$

any (continuous) convex function φ on $[0, 1]$ can be uniformly approximated by a function of the form (*Continued on next page*)

THEOREM 14.1. *When* $m = 2$, $(\varphi) \Leftrightarrow (F)$.

Then by Theorems 12.1 and 14.1, we have

THEOREM 14.2. *When* $m = 2$, $(\Pi) \Leftrightarrow (F)$.

Referring to the definition of points π_j, π'_k in S^{m-1}, we shall introduce the following condition for $m = 2$.

Condition (c): *All points* π'_k *of* Π' *lie between two consecutive points of* Π. (See Figure 5a).

On Figure 6a, the graphs of the functions $G_Y(t)$ and $G_{Y'}(t)$ are drawn on the same interval [0, 1]. The interval [0, 1] is partitioned into sub-intervals by points $\pi_{11}, \cdots, \pi_{1n}$ and $\pi'_{11}, \cdots, \pi'_{1n'}$. Among these sub-intervals, the ones where

$$G_Y(t) > G_{Y'}(t) \tag{14.5}$$

shall be denoted by $I_+^{(1)}, I_+^{(2)}, \cdots$, and the ones where

$$G_Y(t) < G_{Y'}(t) \tag{14.6}$$

shall be denoted by $I_-^{(1)}, I_-^{(2)}, \cdots$. (Note that $G_Y(t) - G_{Y'}(t)$ and $G_{Y'}(t) - G_Y(t)$ are constant positive numbers on $I_+^{(\alpha)}$ and $I_-^{(\beta)}$ respectively.) The length of those intervals will also be denoted by $I_+^{(\alpha)}, I_-^{(\beta)}$, respectively.

Define the positive real numbers $A^{(1)}, A^{(2)}, \cdots$ and $B^{(1)}, B^{(2)}, \cdots$ by

$$A^{(\alpha)} \equiv I_+^{(\alpha)}(G_Y(t) - G_{Y'}(t)) , \quad t \in I_+^{(\alpha)} , \tag{14.7}$$

$$B^{(\beta)} \equiv I_-^{(\beta)}(G_{Y'}(t) - G_Y(t)) , \quad t \in I_-^{(\beta)} . \tag{14.8}$$

Then by (14.4),

$$A^{(1)} + A^{(2)} + \cdots = B^{(1)} + B^{(2)} + \cdots . \tag{14.9}$$

The following theorem gives a simpler characterization of "greater informativeness" than Blackwell-Girshick's Theorem 12.4.1 [4].

THEOREM 14.3. *When* $m = 2$, $(\Pi) \Leftrightarrow (c)$.

PROOF. Without loss of generality, let $\pi_{11} < \pi_{12} < \cdots < \pi_{1n}$ and $\pi'_{11} < \pi'_{12} <$

$$g(u) = \sum_s c_s f_{t_s}(u) + au + b ,$$

where $c_s \geqq 0$. Now without loss of generality, we assume that

$$\pi_{11} < \pi_{12} < \cdots < \pi_{1n} \quad \text{and} \quad \pi'_{11} < \pi'_{12} < \cdots < \pi'_{1n'} .$$

Then it is clear that we have

$$F_Y(t) = \sum_{j=1}^{n} q_j f_t(\pi_{1j}), \; F_{Y'}(t) = \sum_{k=1}^{n'} g'_k f_t(\pi'_{1k}) .$$

Then

$$\sum_j q_j g(\pi_{1j}) = \sum_s c_s F_Y(t_s) + ar_1 + b ,$$

$$\sum_k q'_k g(\pi'_{1k}) = \sum_s c_s F_{Y'}(t_s) + ar_1 + b ,$$

since

$$\sum q_j \pi_{1j} = \sum q_k \pi'_{1k} = r .$$

Therefore condition (φ) is equivalent to condition (F), since $c_s \geqq 0$.

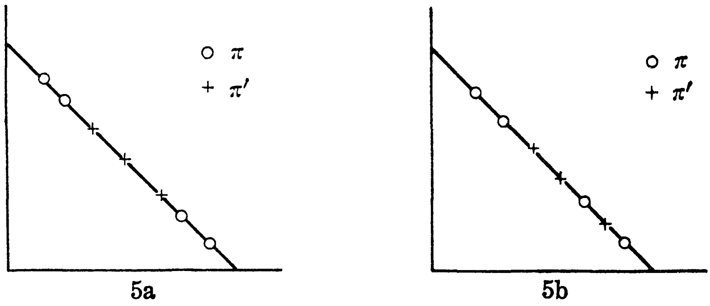


FIGURE 5
DICHOTOMIES: $m = 2$

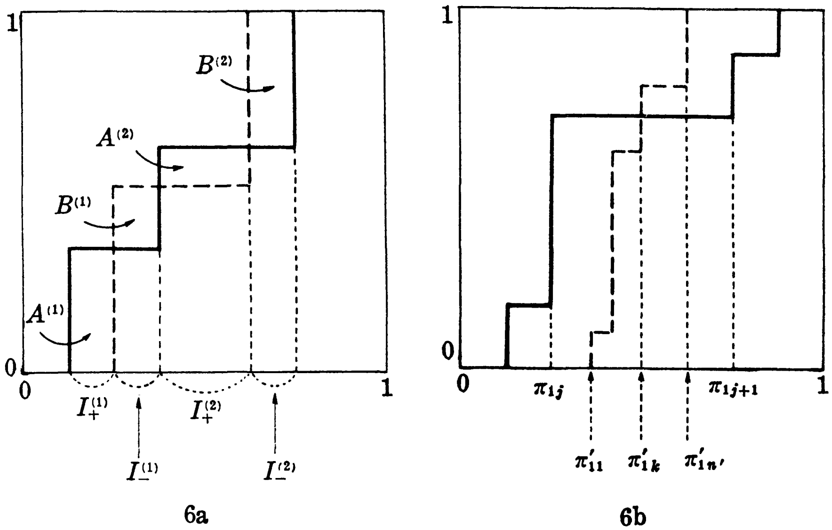


FIGURE 6

$\cdots < \pi'_{1n'}$. First, we shall prove (c) $\Rightarrow$ (II). For a fixed j, we have by condition (c)

(14.10) $$\pi_{1j} \leqq \pi'_{11} < \cdots < \pi'_{1n'} \leqq \pi_{1j+1} .$$

Let q, q' be any vectors such that (14.1) holds. Then $G_Y(\pi_{1j}) < 1 = G_{Y'}(\pi'_{1n'})$. Therefore, there exists an index k such that

(14.11) $$G_{Y'}(\pi'_{1k-1}) \leqq G_Y(\pi_{1j}) \leqq G_{Y'}(\pi'_{1k}) ,$$

(where, when $k = 1$, we consider only the right-hand inequality). Then from the monotonicity of $G_Y(t)$ and $G_{Y'}(t)$ (see Figure 6b)

(14.12) $$\begin{aligned} G_Y(t) &\geqq G_{Y'}(t) , & 0 < t < \pi'_{1k} , \\ G_{Y'}(t) &\leqq G_Y(t) , & \pi'_{1k} \leqq t \leqq 1 . \end{aligned}$$

Therefore, all intervals $I_+^{(1)}, I_+^{(2)}, \cdots$ are located on the left-hand side of π'_{1k}

and all the intervals $I_-^{(1)}, I_-^{(2)}, \cdots$ are located on the right-hand side of π'_{1k}. Therefore by (14.9) and the definitions of $F_Y(t)$ and $F_{Y'}(t)$, clearly, condition (F) holds. Then by Theorem 14.2, it follows that (c) $\Rightarrow$ (Π).

Next, in order to prove that (Π) $\Rightarrow$ (c), suppose that (c) does not hold. Then there exists at least one pair of points among $\pi'_{11}, \cdots, \pi'_{1n'}$, — say π'_{1k} and π'_{1k+1}—such that the interval (π'_{1k}, π'_{1k+1}) contains at least one point of $\{\pi_{11}, \cdots, \pi_{1n}\}$ —say $\pi_{1j}, \cdots, \pi_{1j+s}$, $s \geqq 0$. We shall show that there exists a pair of vectors $q \equiv (q_1, \cdots, q_n)$, $q_j > 0$, $\sum q_j = 1$ and $q' \equiv (q'_1, \cdots, q'_{n'})$, $q'_k > 0$, $\sum q'_k = 1$ such that

$$\Pi q = \Pi' q' \, , \tag{14.13}$$

and for the two corresponding information systems Y and Y', the respective functions $F_Y(t)$ and $F_{Y'}(t)$ do not satisfy condition (F). Then by Theorem 14.2, we can conclude $(\Pi, q) \not> (\Pi', q')$, i.e., (Π) does not hold.

We can choose $q \equiv (q_1, \cdots, q_n)$ and $q' \equiv (q'_1, \cdots, q'_{n'})$ so as to satisfy (14.9) and the following conditions (14.14.1)-(14.14.5):

$$0 < q_g \ll q_h \, , \qquad g = 1, \cdots, j-1, j+1, \cdots, j+s-1, \; j+s+1, \cdots, n, \quad h = j, j+s \, , \tag{14.14.1}$$

$$\sum_{g=1}^{n} q_g = 1 \tag{14.14.2}$$

$$0 < q'_g \ll q'_h \, , \qquad g = 1, \cdots, k-1, k+2, \cdots, n', \; h = k, k+1 \, , \tag{14.14.3}$$

$$\sum_{h=1}^{n'} q'_h = 1 \, , \tag{14.14.4}$$

$$\sum_{g=1}^{j-1} q_g < \sum_{h=1}^{k} q'_h < \sum_{g=1}^{j+s} q_g \, , \tag{14.14.5}$$

where for two positive numbers a and b, $a \ll b$ denote that a is infinitesimally small compared to b. Now, in numbering the intervals $I_+^{(\alpha)}$ and $I_-^{(\beta)}$, let us define $I_+^{(1)}$ and $I_-^{(1)}$ by $I_+^{(1)} \equiv [\pi_{1j+s}, \pi'_{1k+1}]$, $I_-^{(1)} \equiv [\pi'_{1k}, \pi_{1j}]$. Then

$$A^{(1)} \equiv (\pi'_{1k+1} - \pi_{1j+s})(G_Y(t) - G_{Y'}(t)), \qquad \pi_{1j+s} \leqq t \leqq \pi'_{1k+1} \, , \tag{14.15}$$

$$B^{(1)} \equiv (\pi_{1j} - \pi'_{1k})(G_{Y'}(t) - G_Y(t)) \, , \qquad \pi'_{1k} \leqq t \leqq \pi_{1j} \, . \tag{14.16}$$

Then in addition to (14.9), by (14.14.1) and (14.14.3), we have

$$A^{(2)} + A^{(3)} + \cdots \ll A^{(1)} \, , \tag{14.17}$$

$$B^{(2)} + B^{(3)} + \cdots \ll B^{(1)} \, . \tag{14.18}$$

Then by (14.9)

$$A^{(1)} - B^{(1)} = (B^{(2)} + B^{(3)} + \cdots) - (A^{(2)} + A^{(3)} + \cdots) \, . \tag{14.19}$$

Therefore, by (14.17), (14.18), $A^{(1)} - B^{(1)}$ is almost zero. Accordingly by (14.17)

$$A^{(2)} + A^{(3)} + \cdots \ll B^{(1)} \, . \tag{14.20}$$

This implies that we cannot have $F_Y(t) \geqq F_{Y'}(t)$ over the whole sub-interval (π'_{1k}, π_{1j}). *We have thus shown that if (c) does not hold, then (F) does not hold and by Theorem 14.2, (Π) cannot hold. Hence (Π) $\Rightarrow$ (c), completing the* proof of the theorem.

15. SUMMARY

Given the set X of states of nature, described in unlimited detail and with a probability measure defined on it, we consider its partitions: Z (set of events z) and $Y, Y', Y'', \cdots$ (information systems, i.e., sets of mutually exclusive messages denoted by y in Y, y' in Y' etc.). The set Ω_Z consists of all payoff function ω for which Z is a payoff-adequate (i.e., not too coarse) partition of X. Let the set of feasible decisions d entering the argument of ω be D^ω; thus ω is a real valued function on $Z \times D^\omega$. The (gross) expected payoff of an information system Y is

$$U(Y; \omega) = \sum_{z,y} \Pr(x \in z \cap y) \cdot \omega(z, d_y) ,$$

where d_y is a decision responding optimally to message y; that is,

$$\sum_{z} \Pr(x \in z \cap y)[\omega(z, d_y) - \omega(z, d)] \geqq 0 , \qquad \text{all } d \in D^\omega .$$

A class of partially ordering relations is defined on the set of information systems $Y, Y', \cdots$, as follows: Y is said to be more informative than Y', as revealed by a given property of the distribution on $Z \times Y \times Y'$, if this property implies that

$$U(Y; \omega) \geqq U(Y'; \omega) , \qquad \text{all } \omega \in \Omega_Z .$$

In particular, greater informativeness of Y relative to Y' may be revealed by comparing the (bivariate) distributions on $Z \times Y$ and on $Z \times Y'$; by comparing the matrix of likelihoods $\Lambda \equiv [\lambda_{zy}] \equiv [\Pr(x \in z \cap y)/\Pr(x \in z)]$, with the correspondingly defined matrix Λ'; or by comparing the matrix of posterior probabilities, $\Pi \equiv [\pi_{zy}] \equiv [\Pr(x \in z \cap y)/\Pr(x \in y)]$, with the correspondingly defined matrix Π. The latter comparison is the strictly stronger one (Theorem 5.3) except in the important case when the messages of the more informative system (each message being represented by a column of Π, or of Λ) are linearly independent: Theorems 9.5, 9.7. The numerical criteria for comparison of informativeness, given in Theorem 13.2 are all equivalent, and are strictly stronger than that of Theorem 9.1; but they are all equivalent in the case of linear independence between messages. This case applies in particular to information systems that are noiseless with respect to—i.e., are sub-partitions of—the set Z of events (Section 11); or are binomial—the case of two messages,—or binary—the case of two messages and two events (Section 10). Furthermore, for the "dichotomy" cases (i.e., the case of only two events), a certain property of posterior probabilities is proved to be necessary and sufficient for Y to be more informative than Y' (Section 14). The "equivocation" parameter of classical information theory, as well as its generalization to any other convex operator on posterior probabilities, is shown to be a necessary but not sufficient criterion of comparative informativeness; but a stronger criterion involving *all* convex operators on posterior probabilities is shown to be both sufficient and necessary (Section 12).

Statements on comparative informativeness can also be derived from the knowledge, not of the numerical properties of the probability distribution

of events and messages, but of the process by which messages constituting one information system are generated from those of another system. Thus it is shown that "garbling," of Y into Y', a well-defined property of some distributions on $Z \times Y \times Y'$, can never make Y' more informative than Y. A special case of garbling is "collapsing" information, and a still more special case is dis-"joining" it: Theorem 7.3. Moreover, in the case of garbling of linearly independent messages, the Markov matrices used in some of the criteria of informativeness and relating the distribution on $Z \times Y$ to that on $Z \times Y'$ receive an insightful interpretation in terms of conditional probabilities of the messages of one system, given the message of another: Theorem 9.3.

To use our results for a definitive economic comparison of information systems one would have to know information costs. Unless these are additive, one would have to discard the concept of gross payoff functions (Section 4). One should, moreover, characterize the feasible set Δ_φ of decision rules (rather than the set D^φ of feasible decision), and take account of the cost which is associated with a decision rule; this cost, too, is in general non-additive, thus calling for a further re-definition of (net) payoffs.

In addition to these open questions of an economic nature, the mathematician will notice that by confining ourselves to finite sets of events and messages, we have neglected measure-theoretical difficulties which were attacked by other authors (see footnote 14).

More general studies are therefore necessary. We hope the present study contributes to stimulating and preparing them.

University of California, Los Angeles, U.S.A.
and University of Tokyo, Japan

REFERENCES

[1] ABRAMSON, NORMAN, *Information Theory and Coding* (New York: McGraw-Hill, 1963).

[2] BLACKWELL, DAVID, "Comparison of Experiments," in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley: University of California Press, 1951), 93-102.

[3] ———, "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, XXIV (June, 1953), 265-73.

[4] ——— and M. A. GIRSHICK, *Theory of Games and Statistical Decisions* (New York: John Wiley and Sons, 1954).

[5] BOHNENBLUST, H. F., L. S. SHAPLEY and S. SHERMAN, "Reconnaissance in Game Theory," Research Memorandum RM-208 (The RAND Corporation, 1949).

[6] DEGROOT, M. H., "Uncertainty, Information and Sequential Experiments," *Annals of Mathematical Statistics*, XXXII (June, 1962), 602-5.

[7] FEINSTEIN, AMIEL, *Foundations of Information Theory* (New York: McGraw-Hill, 1958).

[8] HALMOS, PAUL R., *Measure Theory* (Princeton, N.J.: Van Nostrand, 1950).

[9] KARLIN, SAMUEL, *Mathematical Methods and Theory in Games, Programming, and Economics*, Vol. II (Reading, Mass.: Addison-Wesley, 1959).

[10] MARSCHAK, JACOB, "Remarks on the Economics of Information," in *Contributions to Scientific Research in Management* (printed privately).

[11] ———, "The Payoff-Relevant Description of States and Acts," *Econometrica*, XXXI (October, 1963), 719-29.

[12] ——— and R. RADNER, "Economic Theory of Teams," Chapter II, Working Paper, No. 82 (Center for Research in Management Science, University of California, Berkeley, 1959).

[13] MCGUIRE, C. B., "Comparisons of Information Systems," Cowles Foundation Discussion Paper No. 71 (1959).

[14] MIYASAWA, KOICHI, "Information Structures in Adaptive Programming," Working Paper No. 61 (Western Management Science Institute, University of California, Los Angeles, 1964).

[15] RAIFFA, HOWARD and R. SCHLAIFER, *Applied Statistical Decision Theory* (Cambridge, Mass.: Harvard University Press, 1961).

[16] SAVAGE, LEONARD J., *The Foundations of Statistics* (New York: John Wiley and Sons, 1954).